1

AD-A155 876

Low Cost Voice Processor

Texas Instruments Incorporated
Dallas, Texas 75265

March 1980

Semi-Annual Report for Period 24 July 1979 - 31 January 1980

APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED (A)

Sponsored by
Defense Advanced Research Projects
Agency under ARPA Order No.: 3335.
Monitored by Dr. Jos. Killiany,
Naval Research Laboratory, under
Contract No. N00173-79-C-0224

DTIC FILE COPY

DTIC
S ELECTE D
JUN 1 8 1985
G

85 06 13 023

SEMI-ANNUAL REPORT
CONTRACT NO. N00173-79-C-0224
24 July 1979 - 31 January 1980

ARPA Order Number:	3335
Program Code Number:	
Name of Contractor:	Texas Instruments Incorporated Central Research Laboratories 13500 N. Central Expressway P. O. Box 225936, M. S. 134 Dallas, Texas 75265
Effective Date of Contract:	24 July 1979
Contract Expiration Date:	31 July 1981
Contract Number:	N00173-79-C-0224
Principal Investigators:	R. K. Hester (214) 238-2367 B. G. Secrest (214) 980-6181
Short Title of Work:	Low Cost Voice Processor
Amount of Contract:	$565,299.00
Contract Period Covered by Report:	24 July 1979 - 31 January 1980

This research was sponsored by the Defense Advanced Research Projects Agency under ARPA Order No.: 3335 Monitored by Dr. Jos. Killiany, Naval Research Laboratory, under Contract No. N00173-79-C-0224

Accession For
NTIS GRA&I ☒
DTIC TAB □
Unannounced □
Justification

By
Distribution/
Availability Codes
Dist | Avail and/or Special
A

DTIC
COPY
INSPECTED
2

Section I
Introduction

This report discusses progress on the development of a potentially low-cost, highly integrated vocoder based on the Belgard algorithm. Work is reported in two areas: The continuing development of the channel bank analyzer and synthesizer integrated circuits and the initial work to develop a pitch tracking chip. These two areas of research are described in Section II and Section III, respectively. Each section is, for the most part, self-contained, but because Section II addresses only changes in the Belgard ICs, the final report for DARPA Contract No. N00173-77-C-0100, published in January 1979, may be referred to for a complete picture of the present design.

Section II

Channel Bank Vocoder Integrated Circuit Development

This section describes the details of the first IC redesign; the subsequent evaluation; and the second redesign, just complete at this writing. The descriptions of the analyzer and synthesizer are separate. In each, the history of one circuit block at a time is discussed chronologically, beginning with a summary of the problems of its initial design. This section is not meant to be a complete IC documentation. Detail is given relating only to changes; circuit blocks with acceptable performance on the initial design are not discussed here. Complete documentation will be presented in a future report when test results of the final designs have been obtained.

A. Channel Bank Analyzer

All the problems associated with the bandpass filters in the initial design were associated with the output amplifier (differential charge amplifier, DCI). The DCI had poor common mode rejection that was very sensitive to a required external bias voltage. In addition, the switched capacitor feedback clocks were phased improperly, resulting in a reduced voltage on the CCD sense gates, and hence, a reduced CCD signal capacity. The filters were sufficiently functional to determine that the center frequencies and bandwidths were acceptable.

The filter redesign consisted of replacing the DCI with another differential amplifier circuit that required no external bias and correcting the clock phase in the feedback loop. These redesigned filters performed as predicted, and no further modifications are required. Schematics of the DCI circuit and the feedback clock phase are shown in Figure 1 and 2, respectively.

A simplified schematic of the half-wave rectifier as originally designed is shown in Figure 3. During R1, current flows through the rectifying transistor,

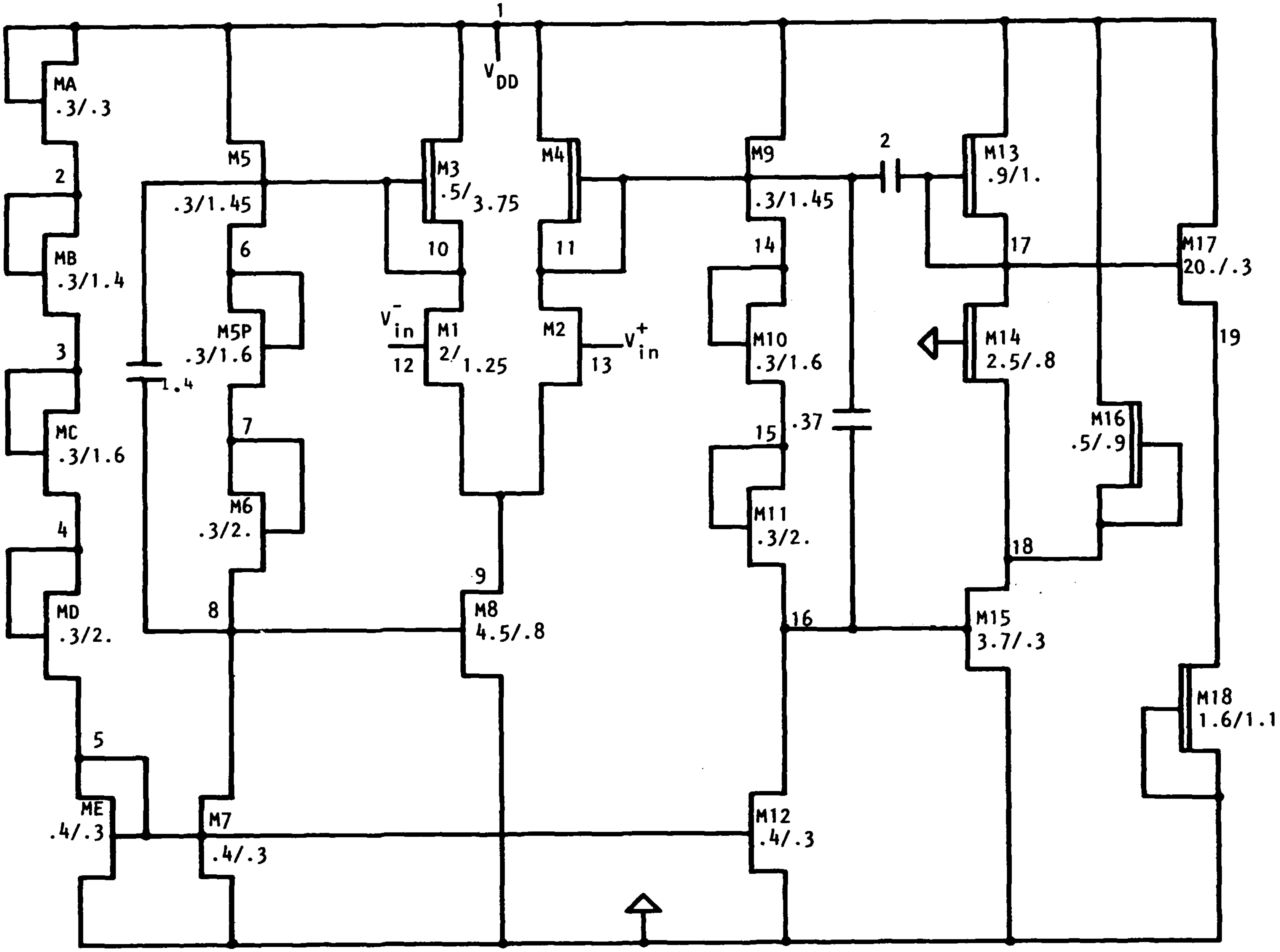

Figure 1 Analyzer Bandpass DCI

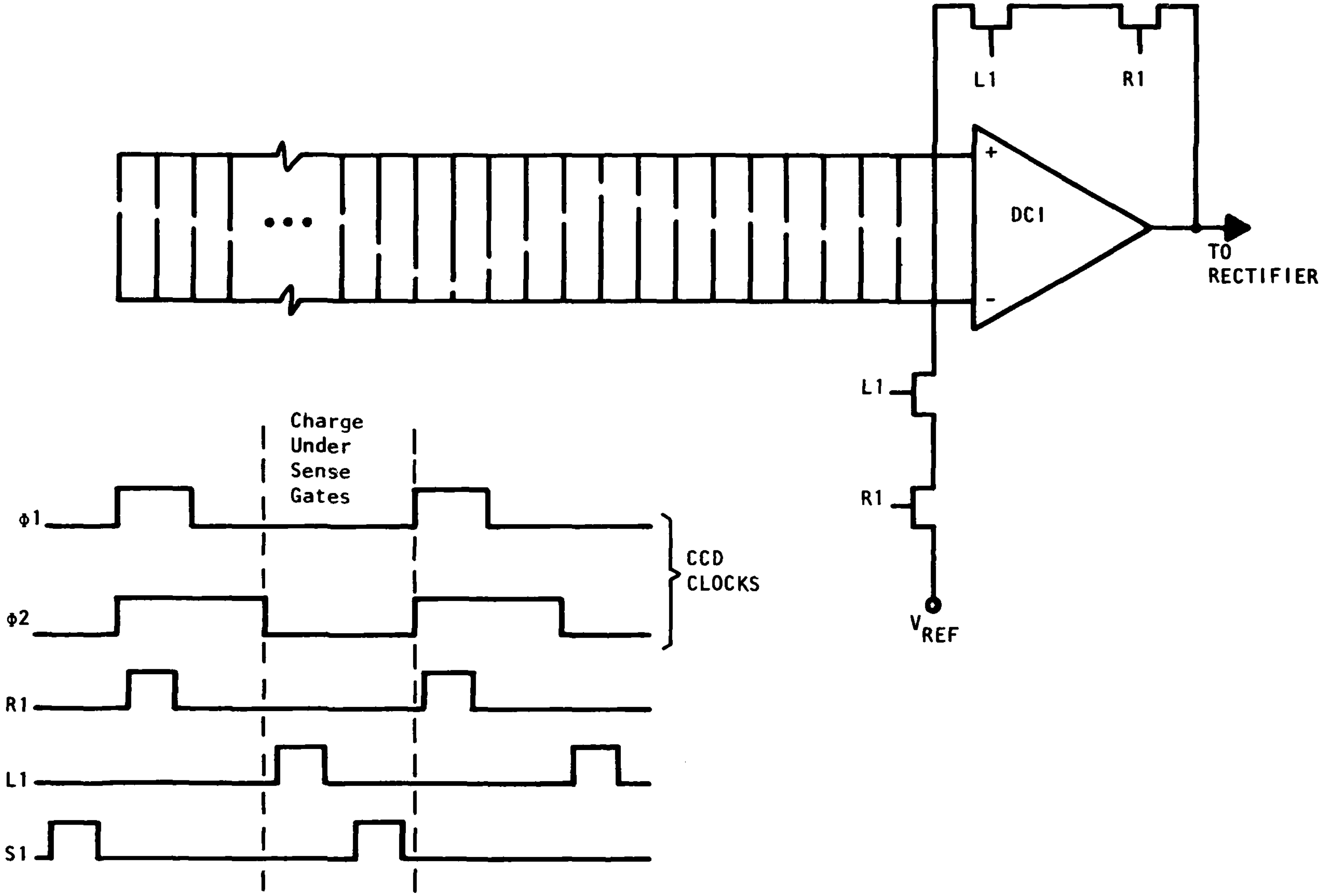

Figure 2 DCI Biasing Technique

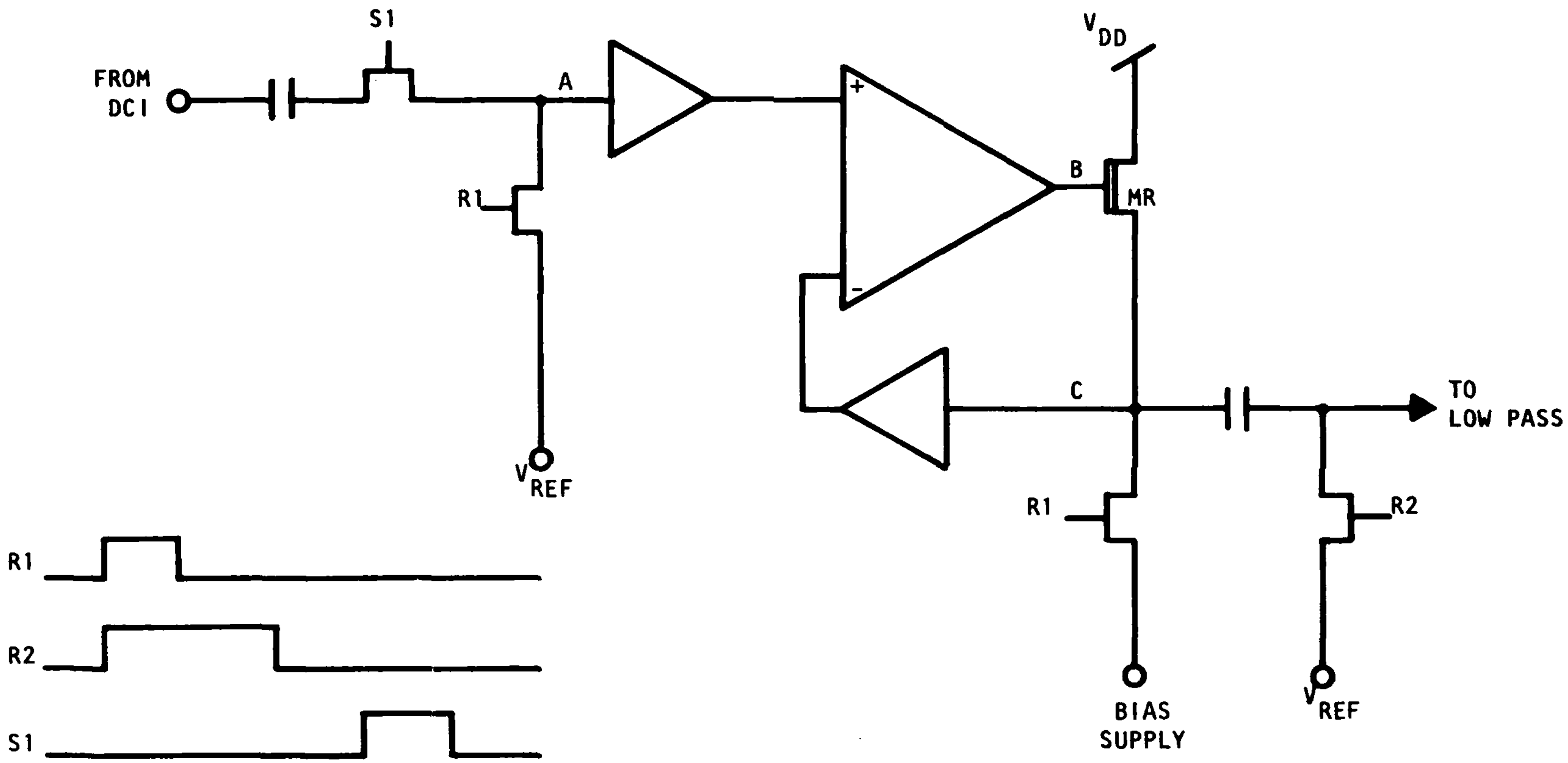

Figure 3 Original Rectifier Design

MR, while the rectifier input is set to V_{Ref}. At the end of R1 the feedback loop should stabilize with MR just at threshold. Any increase of the gate voltage of MR during S1 will raise node C, where a decrease at B leaves C unaffected.

There were two major flaws in the original design. One flaw was that R1 pulled too much current through MR. This resulted in an unstable feedback loop. MR did not settle to its threshold condition after reset, but was turned off by ~ 100 mV instead. This problem prevented input signals with amplitudes less than 100 mV from being rectified, but this effect was masked by the other design flaw. The second problem was that, because the R1 transistor was connected to a low impedance source and the S1 transistor was not, the clock feedthroughs to node A differ. When R1 turned off, node A did not couple down as much as it coupled up when S1 turned on. The rectifier interpreted this as signal and rectified it. As a result, the rectifier output followed any signal greater than ~-400 mV.

A schematic of the redesign rectifier is shown in Figure 4. The extra bias supply was replaced by a current-limiting circuit, MD and ME. These transistors were designed to carry 0.5 µA during R1. However, MD did not behave as predicted, resulting in only 0.2 µA. This current was insufficient to slew node c back to threshold after rectifying large signals. Thus, the rectifier was saturated with 2 V input signals rather than the 5 V design goal. This problem will be overcome in the next version by modifying MD and by connecting its gate to a bond pad so that an external bias can be applied if necessary.

The input structure was modified to equalize the clock coupling from R1 and S1. The R1 transistor was no longer connected to a low impedance source. However there was still a problem due to clock timing. Node A was coupling down before the feedback loop completely settled; consequently, part of the clock coupling was passed to node C, and when S1 turned on, MR began conducting before node A reached V_{Ref}. To correct this problem R1 will be replaced by R2 in the next version. This will allow the feedback loop to settle before node A couples down.

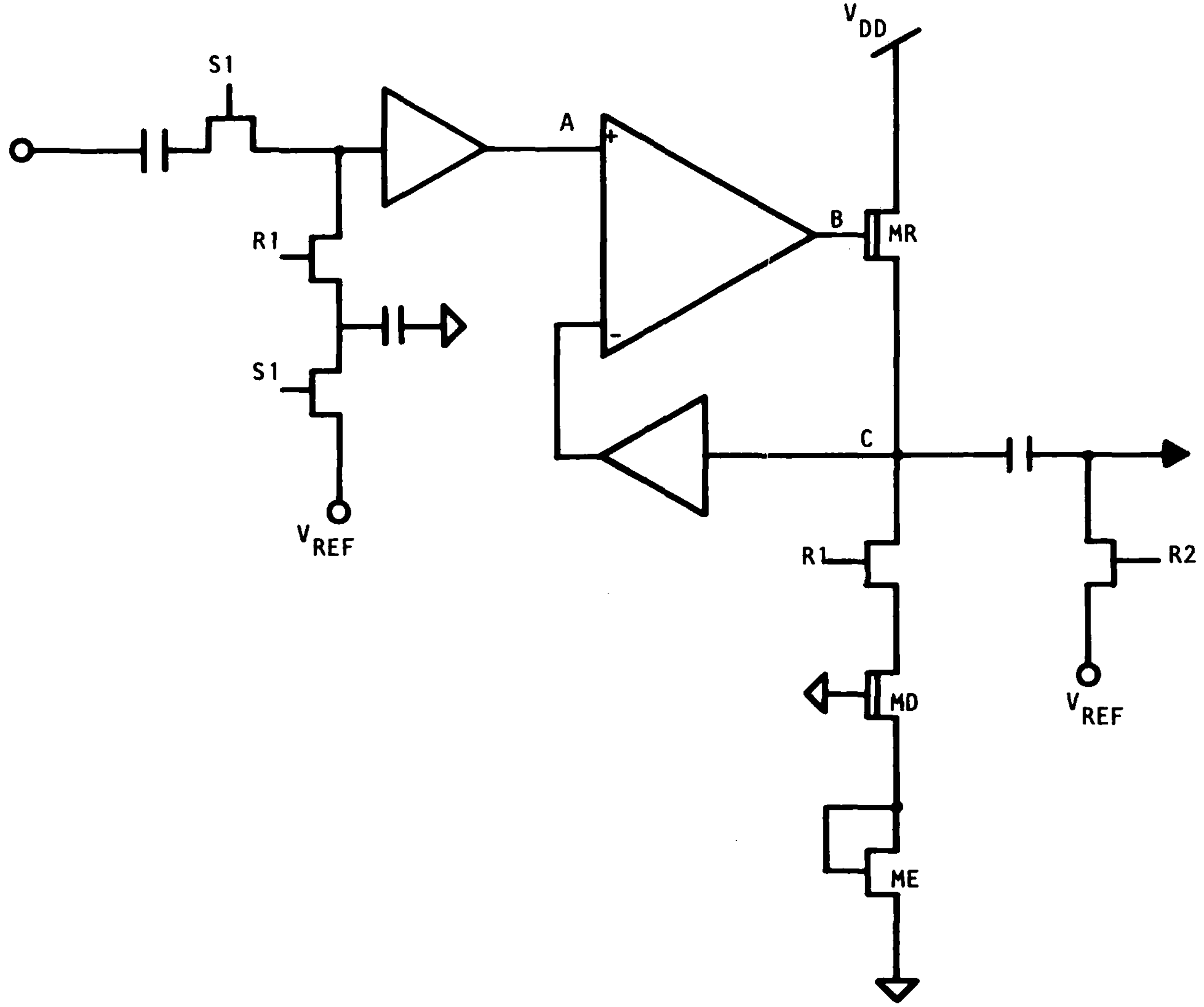

Figure 4 Redesigned Rectifier Topology

The schematic of the lowpass filter as it existed in both the original design and the redesign is shown in Figure 5. Switches S1 and L1 operate a single-pole section while S3 and L2 control a second-order section. Transistor pairs (M21 and M22) and (M23 and M24) are source-follower buffer amplifiers that generate offsets (differences between input and output levels). The scheme originally proposed to eliminate these offsets required clocks R3 and S2. When the reference level is on the lowpass input, S2 turns off and R3 turns on. This passes the reference level through each of the buffers, accumulating the offsets. Clock R4 turns on, storing the offset on the 5.63 pF coupling capacitor. Then R3 turns off and S2 turns on. The problem with the performance of the design is that there is an offset generated by the clocks as well as the buffer amplifiers. In the original design the transistors were unnecessarily large. The redesign used minimum geometry transistors with shield gates to reduce the gate to source capacitance, and hence, the clock coupling. The offset was reduced in the redesign, but because the redesigned analog multiplexer had additional gain, the net effect of the offset was worse in the redesign.

The next version has the modified topology shown in Figure 6. There will now be two lowpass filters in each channel. One will filter the "zero signal" level, and the other will filter the signal. They share buffer amplifiers and are located in very close proximity, so their offsets should be very well matched. The offsets should cancel when the correlated sampling circuit at the output subtracts "zero signal" from the signal filter output.

The price paid for this dual filter approach is twofold. Obviously, the dual filter requires nearly twice as much area as the original design. The other, and equally severe cost is the increased clocking complexity. The clock timing diagram is shown in Figure 7. Twelve clocks are required, in contrast to the six clocks needed in the original design.

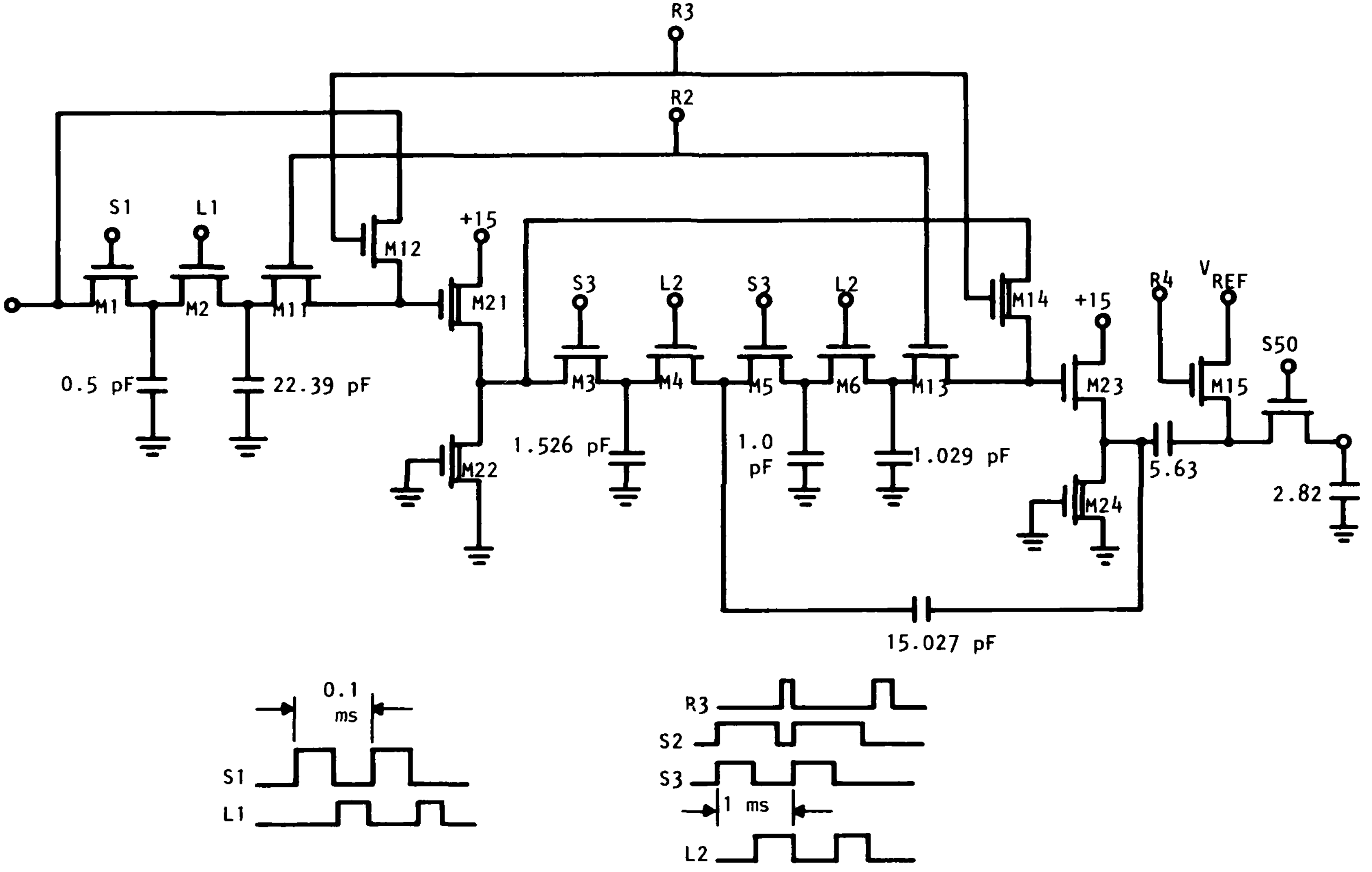

Figure 5 Original Lowpass Design

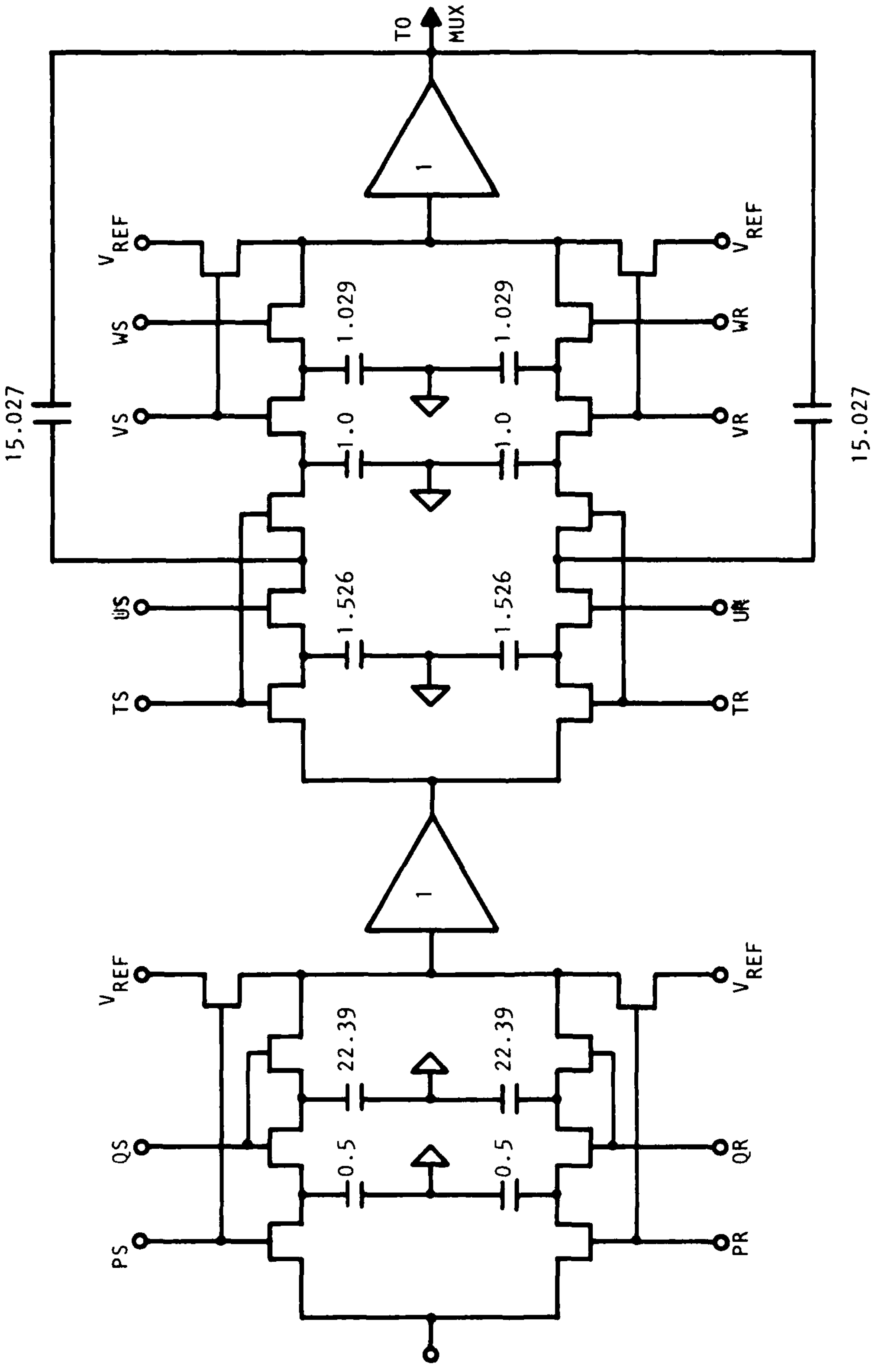

Figure 6 Latest Lowpass Design

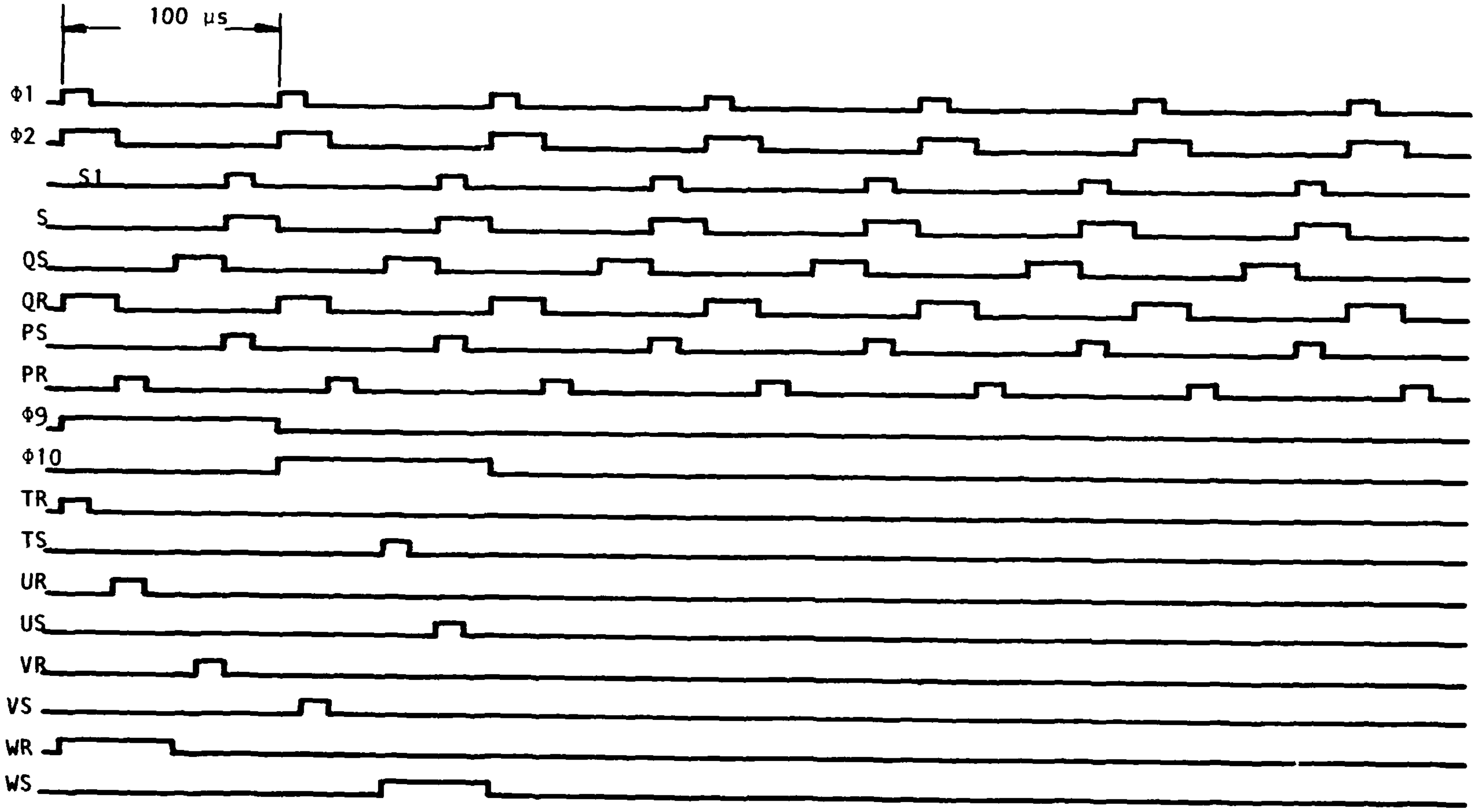

Figure 7 Clock Timing in Most Recent Design

A schematic diagram of the original analog multiplexer design is shown in Figure 8(a). During the lowpass reference period, R4 turned on to store the offset on the coupling capacitor C_c. A 50 Hz exte‚nally supplied synchronization signal generated clock S50 during the signal period of the lowpass. This stored the lowpass output [divided down by $C_c/(C_c + C_s)$] on the storage capacitor C_s. Each of the 19 storage capacitors was then separately selected with an S_j signal for amplification and analog-to-digital conversion. Prior to sampling another channel, the charge associated with the previous channel was eliminated with the $R_{A/D}$ clock.

Several problems were associated with this initial design. Because operational amplifier offsets are unpredictable, the original plan called for a separate, externally adjustable bias supply, MUX REF. This supply was adjusted such that the "zero signal" stored on C_s corresponded to the offset of the amplifier. While this scheme was workable, the separate supply was cumbersome and, in principle, unnecessary. In Figure 8(b) the redesigned multiplexer schematic shows that the lowpass signal with respect to the amplifier offset, not ground, was stored on C_s. This redesigned topology eliminates the extra supply requirement, but inverts the signal. The reinversion was accomplished with a second amplifier that was required to provide additional gain as well. The redesigned multiplexer had 26 dB gain, compared to 13 dB in the original.

Another problem was caused by the asynchronous timing of the S50 clock. In order to minimize the delay time between the external synchronization pulse and the ADC output, the clocks controlling the output were synchronized to the first complete 10 kHz cycle on the chip. This meant that the S50 pulse could occur in any one of ten time positions relative to the 1 kHz clocks. On that one-in-ten occasion when S50 coincided with the S3 clock the outputs of all channels were offset. The redesigned circuit had timing modifications to prevent S50 from coinciding with S3.

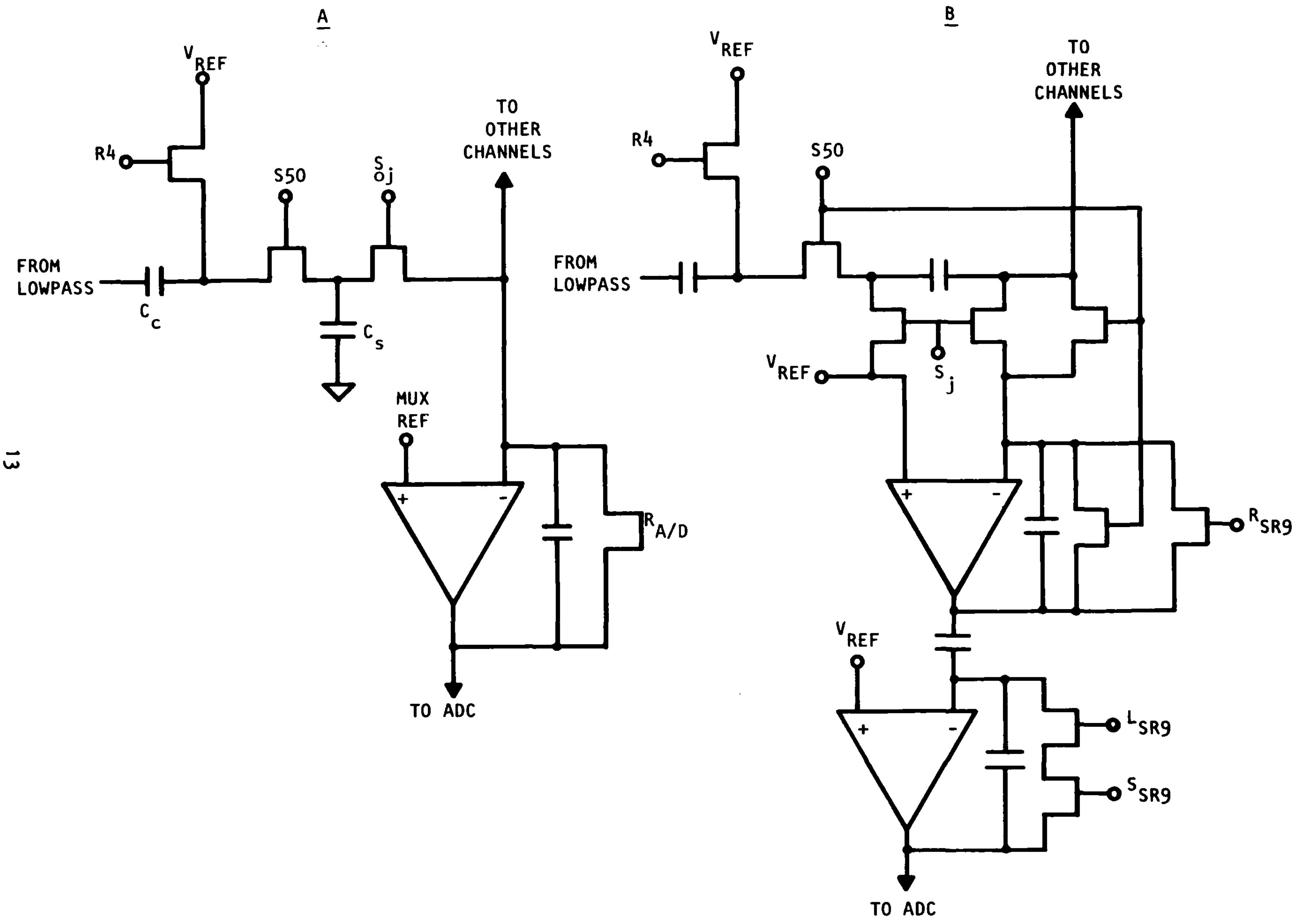

Figure 8 (a) Original Multiplexer Design
(b) First Multiplexer Revision

The only other difficulty associated with the multiplexer was its susceptibility to leakage current. The amplifier inverting input node is connected to a bus line to all 19 channels. The bus had a junction area of nearly 30 mil^2, which produced leakage current discharging the input node and adding a ramp to the signal. The redesign reduced the junction area of the bus to approximately one-quarter of the original area.

The results of the redesign were poor. The increased gain exacerbated each of the problems. The multiplexer was 13 dB more sensitive to leakage current, and since the redesign did not reduce the junction area by that amount, the problem grew worse. The asynchronous, S50-caused offset was also still present and much larger. In addition, elimination of the separate power supply removed the one adjustment that could have made the redesign functional. Not only can the extra supply compensate for the multiplexer offset, but it can also compensate for the average lowpass offset as well. With the additional gain the lowpass offsets were sufficient to saturate the amplifiers. In addition, the reset noise of the first amplifier was amplified in the redesign to an unacceptable level, a problem not encountered in the original design.

A schematic diagram of the newest analog multiplexer is shown in Figure 9. The circuit is very similar to the first redesign, with three major differences. To avoid the asynchronous offset problem, the new design will always sample a frame of speech at 1 kHz unless the external synchronization signal is received, at which time frame sampling ceases until readout is complete in order to avoid skewing the data. The other major differences are the symmetric, fully differential configuration of the amplifiers and the sample-and-hold function performed by L_{SR9} on the reference and $S_{A/D}$ on the signal. Note also that the second gain stage is now reset by S_{SR9} instead of biased with a switched capacitor feedback. This ensures that the reset noise of the first amplifier will be stored on the coupling capacitor and unobserved at the output. The reset noise of the second amplifier still remains, but because its gain is significantly lower, its level should be acceptably low.

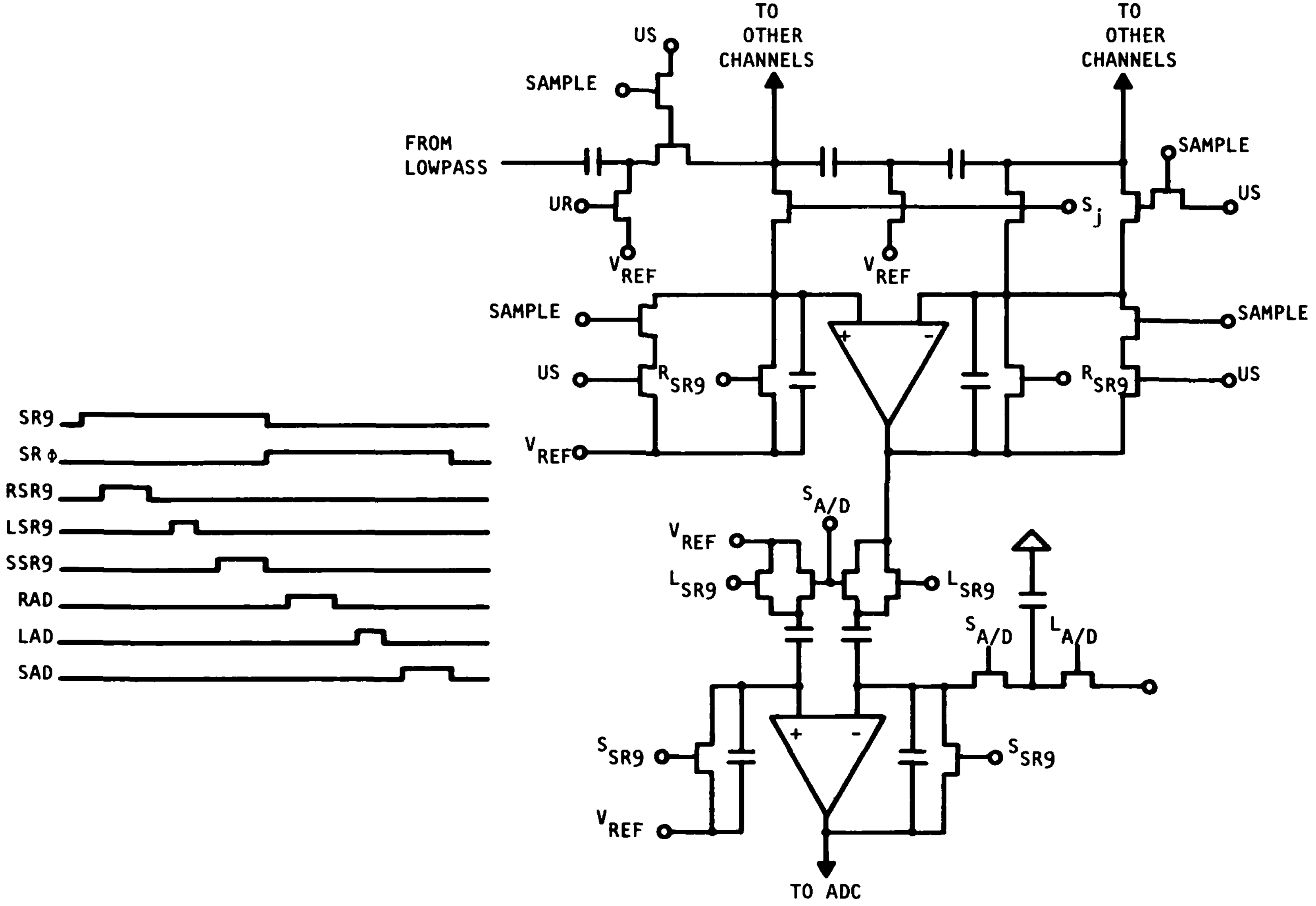

Figure 9 Latest Multiplexer Schematic Diagram

Finally, if all the efforts to remove offsets in the lowpass and multiplexer circuits fail to solve the problems, a separate adjustable bias supply is added to leak through a switched capacitor a small amount of charge to compensate the average offset.

B. Channel Bank Synthesizer

The original syntheizer IC was nonfunctional due to a layout error in the pitch counter which prevented the pitch word from being loaded into the counter. Figure 10 contains a schematic diagram of the two MSB stages of the latest pitch counter. In the original design the FRAME END signal occurred after P1 and before P2. Thus, the input data were replaced by the existing data by the P2 clock before it could be stored during P1. The schematic in Figure 10 indicates the clock timing on the revision. The performance of the revised pitch counter was good except that clock feedthrough from P2 to the floating node (indicated in Figure 10) caused spurious counter load pulses. Decreasing the amplitude of P2 to 12 V reduced the feedthrough enough for the pitch counter to operate correctly. In the more recent redesign the P2 clock was shielded from the floating node by a sheet of polysilicon so that the full 15 V clock amplitudes can be used.

There were two problems in the original design of the noise generator. One was that the rise time of the excitation pulse was too long. This was successfully corrected on the revision by adjusting the sizes of the pullup transistors. The present driver circuit is shown in Figure 11. The other problem in the design was that the voiced/unvoiced decision inhibited the generation of positive noise pulses, but not negative pulses. The revision, shown in Figure 12, worked well. The symbol x represents pitch, positive noise, or negative noise excitation.

The one DAC layout problem resulting in 5 V clock signals in the capacitor array was corrected in the first revision. No further modifications were necessary.

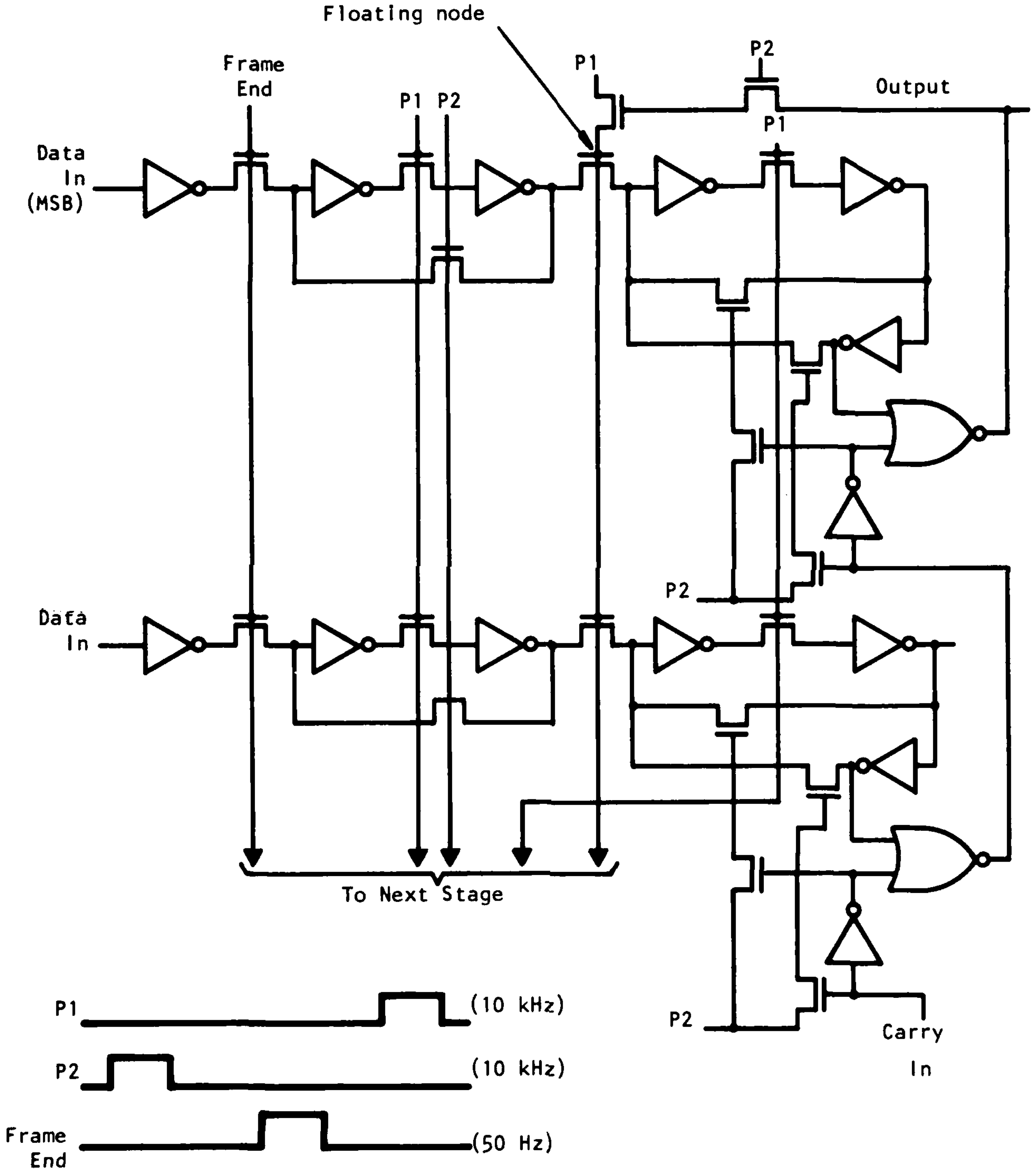

Figure 10 Two MSB Stages of Pitch Counter

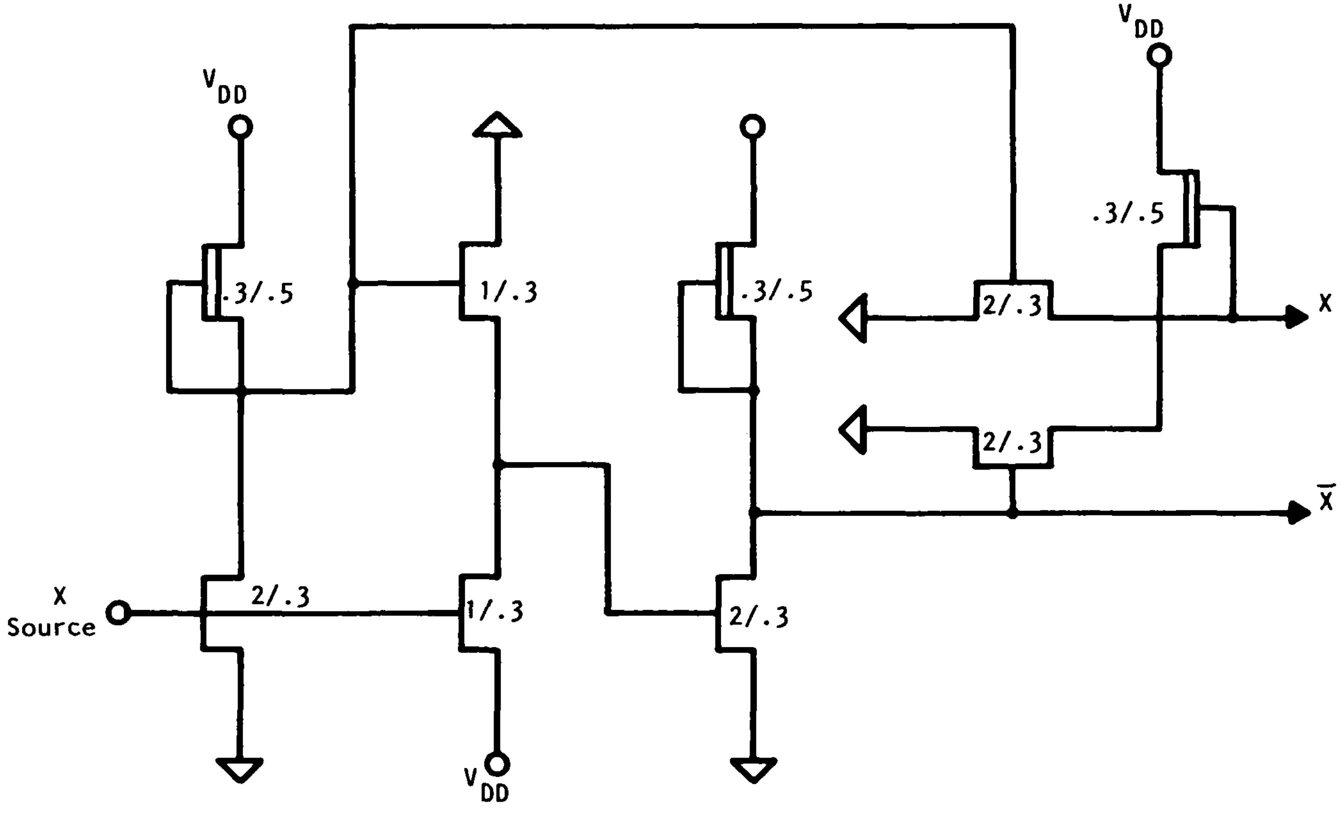

Figure 11 Excitation Driver

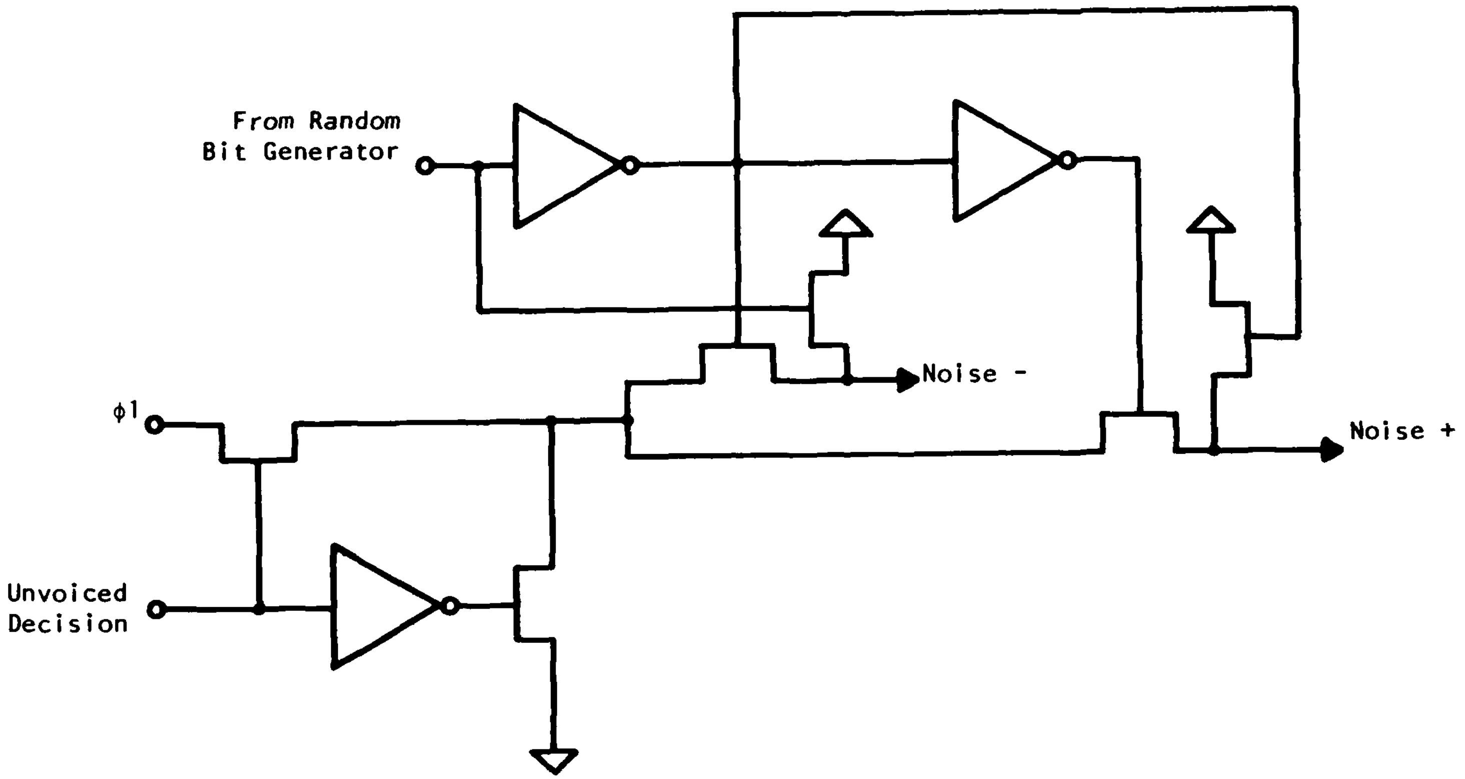

Figure 12 Redesigned Noise Generator

Figure 13 depicts the orignal design of the signal path of a single channel through the bandpass filter. The DAC and unity gain buffers contribute offsets. The modulator is designed to eliminate the total accumulated offset. During R3, the R4 clock stores on the coupling capacitors (C_p, C_{n+}, and C_{n-}) the voltage V_ϕ - V_{Bias1}, where V_ϕ represents the lowpass output during the reference half-cycle. Then, when R3 turns off, the bandpass filter samples V_{Bias2}. Note that the bandpass does not sample during R3. When an excitation pulse occurs (p, n^+, or n^-), the lowpass output is V_{LP}, so the voltage step sampled by the bandpass is $V_{LP} - (V_\phi - V_{Bias1}) - V_{Bias\ 2}$. Let V_Z represent lowpass output corresponding to silence. The bias voltages are externally adjusted such that

$$V_Z - V_\phi = V_{Bias2} - V_{Bias\ 1}.$$

This condition assures that the bandpass samples no excitation during silence, despite the offsets in the signal path.

In the original design a large 50 Hz signal was present at the modulator input when silence was requested. This spurious 50 Hz signal was reduced in magnitude by approximately 26 dB by replacing the single R3 clock by four clocks with staggered falling edges. The clocks were revised to turn off in progression, beginning with the switch closest to the DAC and ending with the reset switch in the second stage of the lowpass. The staggered falling edges were accomplished with simple inverters, and a layout error in one inverter occurred in the revision. (This layout error is corrected in the latest revision.) With the aid of an external inverter the circuit was functional. The revised circuit had a silence offset sufficiently small that it could be eliminated with the two bias supply approach. The remaining problem in the revision was that the silence offsets of the even channels were different from those of the odd channels. This effect is understandable because the even and odd channels are laid out as mirror images. Any misregistration of mask levels during fabrication can cause opposite effects on opposite halves of the chip. In order to provide offset cancellation for all channels simultaneously, another bias supply has been added to the latest design.

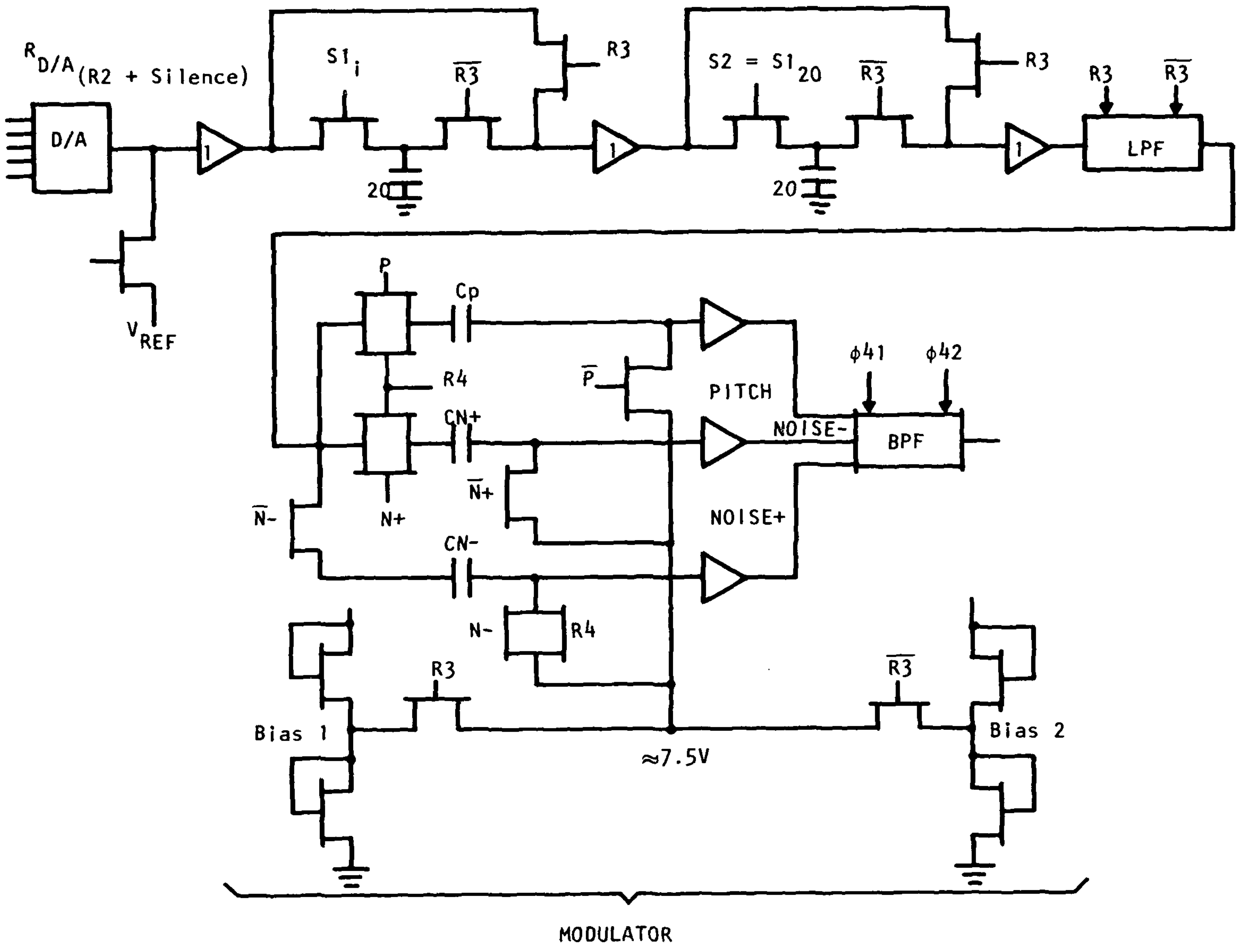

Figure 13 Analog Signal Path of Single Channel

The bandpass filters on the original design had some conceptual as well as layout errors. Layout errors prevented channel 19 from functioning and reduced the gain of channel 16. The filter topology is shown schematically in Figure 14. The capacitor controlling circuit Q, $\alpha_L C_L$, is sensitive to parallel parasitic capacitance. The other problem with the original design was that the capacitor ratios were determined using the S-plane to Z-plane mapping, $S = Z - 1$. Because the filter sampling rate has been insufficiently high, this approximation is not adequate. As a result, the center frequencies and bandwidths significantly differed from those required.

In the revised chip the layout errors were corrected, the capacitor ratios were more accurately chosen, and the filter topology was slightly modified to reduce the performance sensitivity to parasitic capacitance. The revised bandpass is shown in Figure 15. Note the difference in the Q-controlling capacitor configuration and the clock timing change in that same vicinity. A small, but representative, set of performance data is summarized in Table 1. The center frequencies are within 12 Hz of their design goals and the bandwidths within 20 Hz. Notice that there is a systematic increase of bandwidth with increasing frequency. It is thought that this effect is due to capacitor matching errors, and the later revision has attempted to compensate for this problem. The most serious problem in the revised circuit is the low dynamic range, only 30 dB. In the newly revised version 12 dB improvement should be obtained by using lower noise operational amplifiers, and an additional 6 dB increase should be gained by increasing the filter gains at the center frequencies by that amount.

The only other problem with the synthesizer occurred in the summing amplifier. A schematic diagram of the revised circuit is contained in Figure 16. It is a fully differential configuration to sum the even and odd channel signals in opposite phase. However, there are 10 odd channels and only nine even ones. The parasitic capacitance is different on the inverting and noninverting nodes as well. As a result, the odd channels have 2 dB lower gain than the even channels. In the latest design the summing amplifier configuration is symmetric. There should be identical gains for all channels.

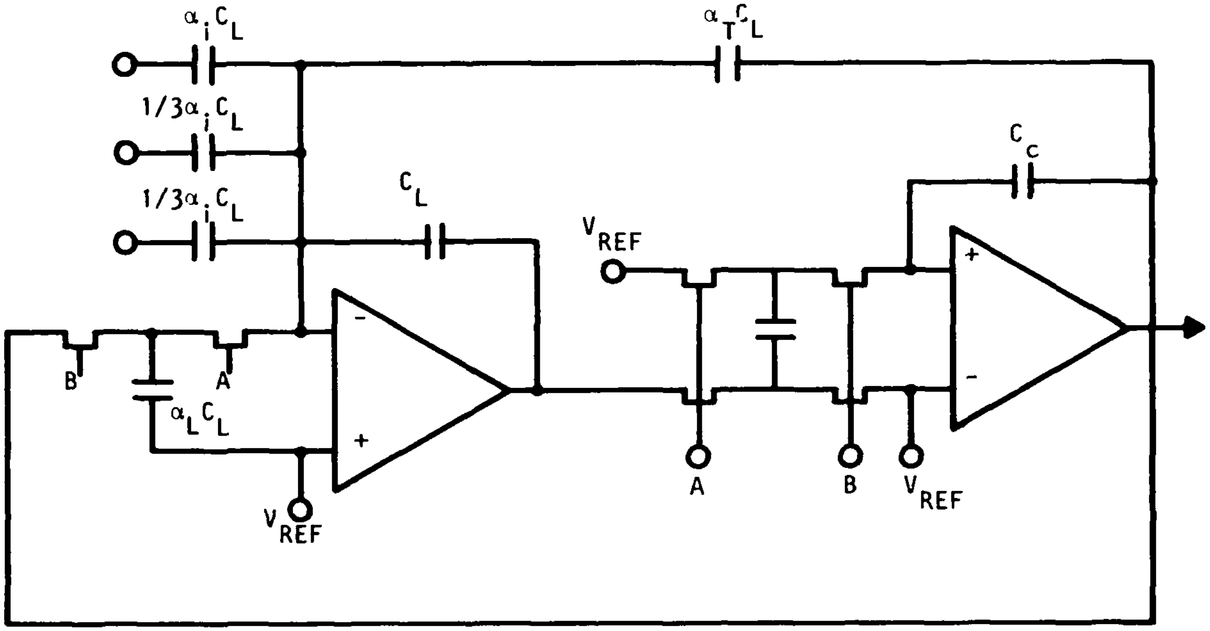

Figure 14 Original Synthesizer Bandpass Design

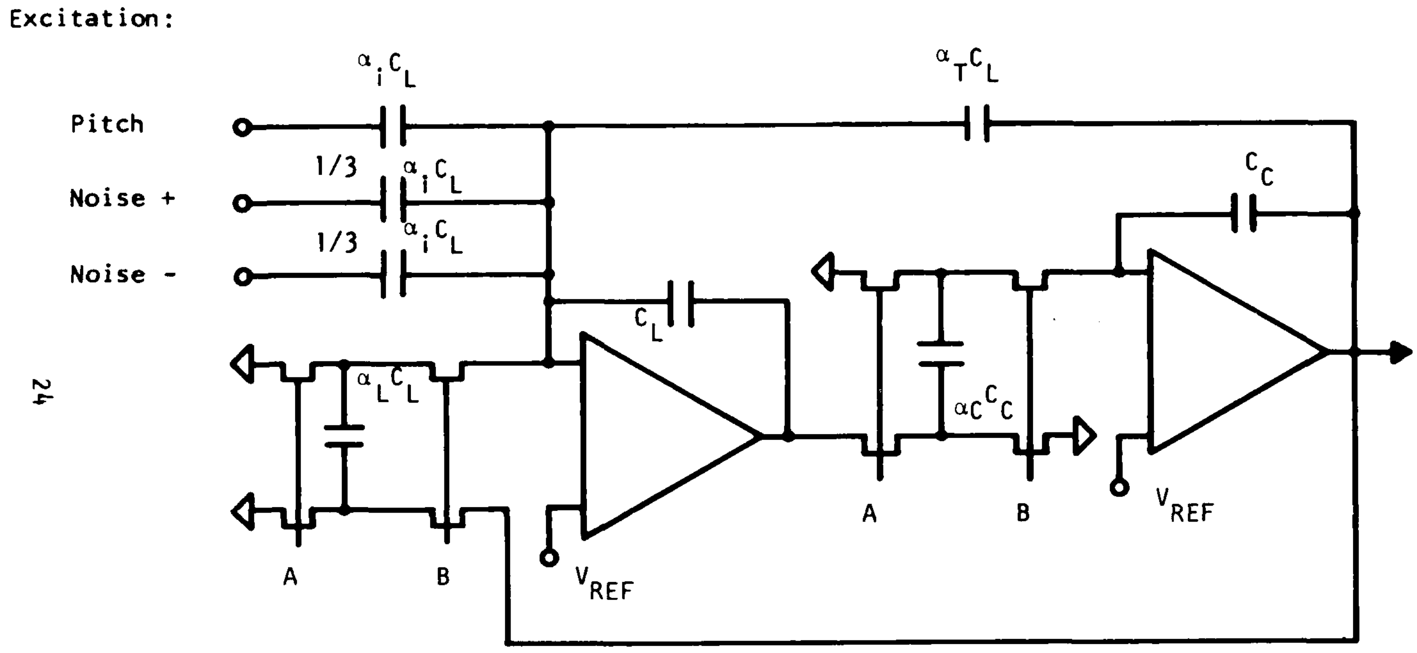

Figure 15 Revised Synthesizer Bandpass

Table 1

Synthesizer Filter Performance Summary

Channel	f_o (Avg.)	σ_{f_o}	BW (Avg.)	σ_{BW}
1	236.2	2.8	46.3	0.16
2	355.9	1.3	47.5	0.21
3	472.5	0.7	48.2	0.14
4	595.3	23.3	49.1	0.86
5	711.0	2.9	50.2	0.49
6	830.8	3.4	50.2	0.26
7	990.1	5.6	52.2	0.11
8	1140.7	4.2	52.6	0.31
9	1288.0	2.0	53.5	0.03
10	1443.2	5.3	54.6	0.14
11	1591.4	4.6	55.1	0.02
12	1794.0	10.2	77.4	0.65
13	1990.4	0.8	77.6	0.16
14	2193.4	9.5	76.9	12.88
15	2386.6	11.0	80.2	0.23
16	2698.3	16.2	83.2	1.12
17	2987.9	7.0	84.8	0.22
18	3301.0	3.4	86.5	0.39
19	3606.6	128.3	-	-

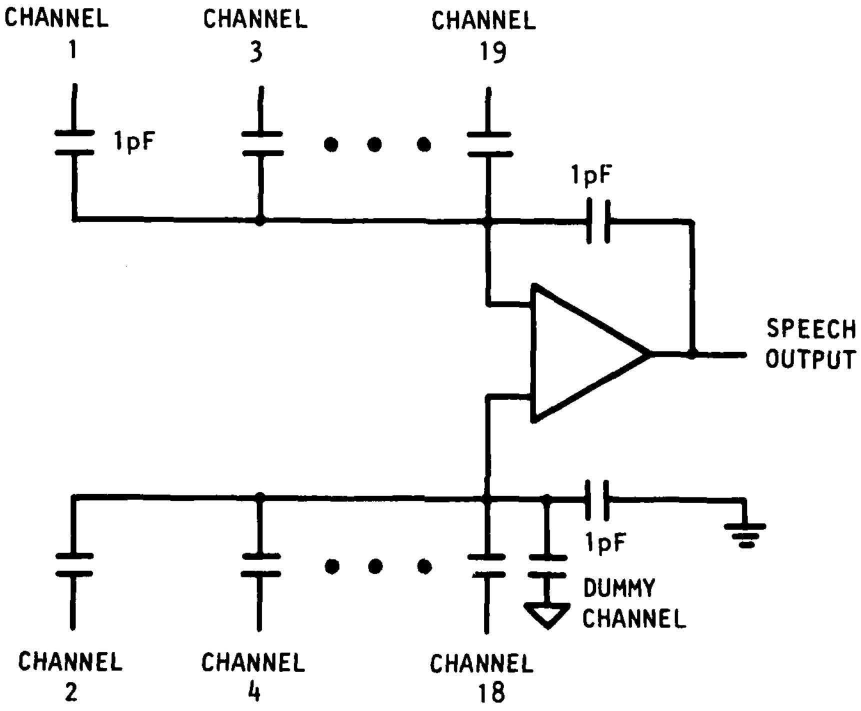

Figure 16 Summing Output Amplifier

Section III

Low Cost Pitch Tracking Development

A. Introduction

A schedule of the major tasks to be performed in the DARPA low cost pitch tracking program is given in Table 2. The following tasks have been completed at this time:

- The data base has been collected, edited, and digitized onto disc.
- Software programs have been written to downsample the data from 12.5 kHz to both 10 kHz and 8 kHz. A software program has been written to add noise to the data.
- A description of the baseline algorithms has been completed and is included for the Gold-Rabiner, harmonic, and correlation pitch trackers. A more complete description of the data base and the baseline algorithms follows in the next two sections.

B. Data base

The pitch base has been collected, digitized, and edited onto disc. The data set consists of 58 speakers, 32 male and 26 female ranging in age from six years to 87 years. Each subject read a series of 11 utterances of approximately three seconds in duration. The last six sentences were constructed to contain approximately 70 within-word phoneme transitions. A complete description of the data base is given in Appendix A.

C. Baseline algorithms

Description of the Gold-Rabiner, harmonic, and correlation pitch trackers are presented. These descriptions represent the baseline algorithms that will provide the base for modifications to achieve an LSI implementation. In each case, an attempt was made to implement the algorithm as presented in the literature, with some modifications to aid in the software simulation.

Table 2

DARPA Low Cost Pitch Tracking Schedule

	1979									1980									1981				
	A	S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M	A	M	J
Establish data base	—	—	—	—																			
Describe baseline algorithms				—	—																		
Develop simulations																							
Harmonic	—	—	—	—	—	—	—	—															
Cepstral	—	—	—	—	—	—	—	—															
Correlation	—	—	—	—	—	—	—	—															
Gold-Rabiner	—	—	—	—	—	—	—	—															
Develop evaluation software						—	—	—															
Algorithm evaluation and iteration						—	—	—	—	—	—	—	—										
Algorithm selection												—											
Simulation of selected algorithm													—	—	—	—	—	—	—	—	—	—	—

- Harmonic Pitch Tracker

The basic algorithm is given in the papers by S. Seneff and P. Bosshart.[1,2] Modifications have been made in the method of obtaining the Fourier transform of the data. The reasons for the modifications are to allow a change in sample rate and to allow a greater range of pitch frequencies to be considered.

Once every FP milliseconds a frame of data will be processed in the following manner (typical value for FP is 10 ms):

A. Preprocess the data.

1. Get a frame of data (WL milliseconds long, typical value for WL is 38 ms).
2. Preemphasize the data. (Preemphasis constant = 0.8.)
3. Calculate the squared energy in the frame of data as:

$$E = \sum_{i=1}^{nx} s_i^2 \quad .$$

where nx = number of samples in frame of data. If the squared energy is less than a fixed threshold, set the frame of speech data to unvoiced (pitch = 0), and return to Step 1. A typical voicing threshold is 10,000.

4. Hamming-window the frame of data using a WL ms window.
5. Take the FFT of the frame of Hamming-windowed data.'

Use an N point transform where N is such that the frequency resolution, DF, is equal to 6.6 Hz, i.e.:

N = 1./(sample period) (DF)

For example, if DF = 6.6 Hz, sample period = 80 microseconds, then N = 1,894.

6. Obtain the magnitude of the spectrum just calculated.

B. Determine the pitch for the frame of data.

1. Find the peaks in the spectrum from F_{min} to F_{max}. (typical values of F_{min} = 210 Hz: F_{max} = 1,100 Hz)

2. Eliminate spurious peaks.

   a. A peak is removed if it is within 6 samples (≈ 40 Hz) of a larger neighboring peak.

   b. A peak that is more than 6 samples, but fewer than 10 samples, from its nearest neighbor is removed if its amplitude is less than one-half amplitude of its nearest neighbor.

3. Rank-order the remaining peaks in descending order of magnitude.

4. Iteratively generate a table of pitch values, with the pitch values in ascending order in each row. First iteration (first row) - enter single pitch value in table (distance between two largest peaks).

   Second iteration (second row) - Third largest peak is added to the list of peaks under consideration - two new pitch estimates are added to the table, defined as the distances between the adjacent peaks. A score is computed for the maximum number of consecutive "equal" pitch estimates in the table, where "equal" is defined as being within two samples (≈ 14 Hz) of the succeeding entry in the table.

   .

   .

   .

   Nth iteration - iterations continue until at least seven "equal" estimates are obtained. If there are fewer than seven "equal" estimates, the iterations continue until the size of the next available leftover peak is less than one-tenth the size of the largest peak, or until a maximum of seven peaks have been exhausted. If either of these conditions is met, iterations stop, even though an inadequate score has been accumulated. Choose the pitch estimate with the best score. In case of a tie, choose the larger pitch estimate.

C. Smoothing and voiced/unvoiced decision.

The above preprocessing and determination of pitch period is done on a frame basis, to obtain a pitch estimate every FP milliseconds. The result is an unsmoothed pitch contour. This pitch contour is now smoothed and a voiced/unvoiced decision is made. This is accomplished by passing the unsmoothed pitch contour through a three-point median smoother, followed by a five-point median smoother.

1. Three-point median smoother

   If none of the three points are "equal," then the frame of speech is unvoiced. Here, "equal" is defined as being within five samples of each other (≈ 33 Hz).

2. Five-point median smoother

   This smoother uses as input the output of the three-point median smoother. If no more than two of the five input samples are "equal," the frame is unvoiced. Here, "equal" means within three samples of each other (≈ 20 Hz).

As can be seen by the above procedure, there will be at least a three-frame lag in outputting the pitch value for the frame of data.

- Gold-Rabiner Pitch Tracker

The basic Gold-Rabiner algorithm is given in the book Theory and Application of Digital Signal Processing, by Gold and Rabiner.[3] The algorithm which has been implemented follows the outline given in the paper by Marilyn Malpass, "The Gold-Rabiner Pitch Detector in a Real Time Environment,"[4] except that some of the thresholds have been made functions of the sample period.

The Gold-Rabiner algorithm works on a frame-by-frame basis:

1. Obtain a frame of speech (typically 10 milliseconds).
2. Low Pass Filter (LPF) the frame of speech (≈ 1,000 Hz); a three-pole Chebychev filter was used.

3. Find the maximum and minimum values of the filtered speech within the frame and check the difference against an energy threshold (= 50). If the energy is low, set frame to unvoiced and set the periods and number of samples since the last successful peak to the intial values in each of the six channels (Table 3). If the energy is above the threshold, perform the peak search.
4. Search the frame of data sample-by-sample for peaks. When a change in slope occurs, take the previous sample as the peak. If it is a negative peak, complement the value (this result may be negative if the value of the peak is positive). Store the peak value as the current positive or negative peak and take the measurements described below. After each sample, decrement the blanking count if greater than zero, increment the number of samples since the last success, and update the current measurement threshold (threshold = old threshold times decay factor), if the blanking count has reached zero. Do this for each of the six channel information blocks, and return to the peak search. Do this for each of the channel information blocks that is affected by the peak just found, and return to the peak search.
5. Take measurements M1, M2, M3 (positive peak) or M4, M5, M6 (negative peak) and store in respective channel blocks: M1, M4: peak value = current positive or negative peak. M2, M5: peak-valley = current positive (negative) peak plus previous negative (positive) peak. M3, M6: peak-peak = current peak peak value minus previous peak value. (See Figure 17.)
Check each of the three measurements as follows:
If the blanking count is not equal to zero or the measurement is less than the threshold, call the measurement a failure and proceed to the next measurement. If the measurement is a

Table 3

Initial Channel Information For Each Of The Six Pitch Detectors

P_A = 20

P_B = 30

P_C = 40

Previous Measurement = 0 (used for M1 and M4)

P_{AV} = 5 milliseconds

Blanking Count = 1 millisecond

Current Measurement Threshold = 0

Decay Factor = Exp $(-0.695/P_{AV})$

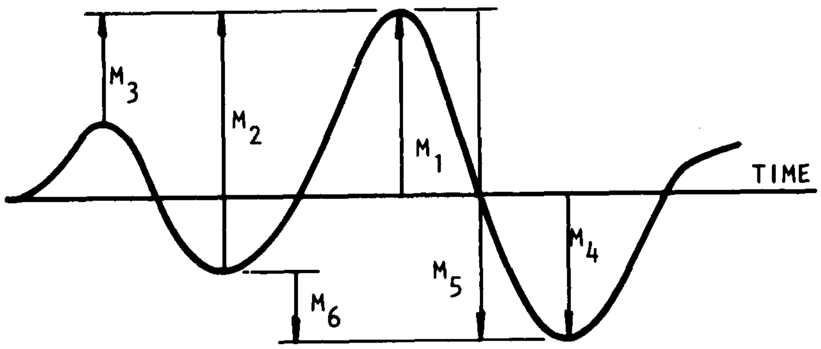

Figure 17 Measured Parameters of Filtered Speech

success, store it as the new threshold, slide periods P_A and P_S to periods P_B and P_C, and store the number of samples since the last success as period P_A. If the previous frame was unvoiced, do not change P average (P_{AV}). If the previous frame was voiced, compute a new P_{AV} = (old P_{AV} + P_A)/2 and confine P_{AV} to be between $P_{AV\ min}$ and $P_{AV\ max}$. ($P_{AV\ min}$ = 4 milliseconds and $P_{AV\ max}$ = 10 milliseconds). Compute the blanking count = 0.4 (P_{AV}), store the appropriate decay factor, and set the number of samples since the last success to zero. [Here decay factor = exp ($-0.695/P_{AV}$)]. Proceed to the next measurement.

6. At the end of the frame of speech data, form a table of 36 pitch periods by storing P_A, P_B, P_C, $P_A + P_B$, $P_B + P_C$, and $P_A + P_C$ from each of the channel information blocks. The six pitch period candidates are the most recent periods, P_A, from the six channels. The pitch period of each candidate being tested determines the window of tolerance. This window is a function of the sampling period, Table 4. A window has four "panes" with associated biases. Each pitch period candidate, P_K, is compared to all 36 values four times as follows:

   a.) Clear pitch period score (PSCORE) for this candidate.

   b.) Clear score counter (SCORE).

   c.) Determine "pane" for pitch period candidate.

   d.) Compare pitch period candidate against all 36 values in table, if $| P_K - P_N | \leq \text{Pane}_K$, increment SCORE, n = 1,...36.

   e.) Add bias for this window pane to SCORE.

   f.) Compute NEW SCORE = SCORE - THRESHOLD (THRESHOLD = 13)

Table 4

Windows Of Tolerance And Bias

		Panes				
Pitch	16-31	1	2	3	4	Window 1
Period	32-63	2	4	6	8	Window 2
Ranges	64-127	4	8	12	16	Window 3
(Milliseconds)	128-255	8	16	24	32	Window 4
Pane Bias		8	6	3	1	

g. Compare magnitudes of NEW SCORE and PSCORE.
   If | NEW SCORE | > | PSCORE |, replace PSCORE with NEW SCORE.

h. Repeat stops b. through g. with remaining "panes."

i. Save PSCORE for this pitch period candidate.

j. Repeat steps a. through i. for each of the remaining pitch period candidates.

7. Pick the winning pitch period from the six candidates by choosing the highest score, PSCORE. If the winning score is negative or if the winning score is greater than P_{max}, set the voiced/unvoiced indicator to unvoiced. (P_{max} = 25.5 milliseconds). If the winning pitch period is accepted, set the voiced/unvoiced indicator to voiced.

- Optimized Correlation Pitch Tracker

The optimized correlation pitch tracker is a pitch tracking algorithm developed at Texas Instruments by George Doddington and Bruce Secrest and is described in the internal report "Optimized Correlation Pitch Tracker for Speech Systems Applications," dated February 1979.

The algorithm consists of three basic parts. First a correlation technique is used to obtain the periodicities of the speech. Then dynamic programming techniques are used to preserve continuity of the pitch track; finally, pattern matching is used to obtain the voiced/unvoiced decision. Since the algorithm has not appeared yet in the published literature, the internal technical report is included as Appendix B.

Section IV

Summary

There is a high confidence level that the recent revision of the Belgard chips will perform adequately. The latest designs have been completed and submitted to the prototype photomask shop. Both sets of masks should be available by 1 March 1980. Processing will proceed simultaneously in both the Advanced Frontend Processing Center (AFPC) and the Central Research Laboratories (CRL). This decision was made so that a processing problem in either facility will not delay fabrication. An optimistic prediction of turnaround in the AFPC is five weeks. In CRL it is a little longer. Turnaround in either facility will not exceed seven weeks under normal circumstances. This means that devices will be under test sometime in the first half of April.

The pitch tracking baseline algorithm simulations are complete, as is the data base to be used in performance evaluation. Work is now being concentrated on the evaluation technique itself. As the redesign of the channel bank chips is now complete, the iteration of pitch tracking algorithms, trading algorithm complexity for implementation ease, will begin. Four algorithms will be examined for integrated circuit implementation in the next six months, with one design emerging as the best candidate for IC implementation.

References

1. S. Seneff, "Real Time Harmonic Pitch Detector," Technical Note 1977-5, Lincoln Laboratory, January 1977.

2. Patrick Bosshart, "Preliminary Study of a Harmonic Pitch Detector Implementation in MOS/LSI," Lincoln Laboratory, May 1978.

3. L. R. Rabiner and B. Gold, Theory and Application of Digital Signal Processing (Prentice-Hall, New Jersey, 1975).

4. M. L. Malpass, "The Gold-Rabiner Pitch Detector in a Real Time Environment," Proceedings of EASCON 1975.

APPENDIX A

DESCRIPTION FOR DATA BASE PITCH

Appendix A

```
DESCRIPTION FOR DATA BASE     PITCH
-----------------------------------

TITLE: PITCH

DIRECTORY NAME: [SPCH2.PITCH]

COLLECTOR: BRUCE SECREST

ROOM DESCRIPTION: TRACOUSTICS RE-244B DOUBLE WALLED SOUND BOOTH

MICROPHONE TYPE: ELECTRO-VOICE RE16; DYNAMIC, CARDIOID

FILE FORMAT DECODING:
   A)CHARS 1-3 : SPEAKERS INITIALS
   B)CHAR  4   : PHRASE NUMBER (1,2,...,9,A,B)
   C)CHAR  5   : SEX (M OR F)
   D)CHAR  6   : AGE GROUP (1 IS <13; 2 IS 13-20; 3 IS 21-39;
                            4 IS 40-69; 5 IS >69)
   F)CHAR  7   : SAMPLE RATE (1 IS 12.5 KHZ, 2 IS 10 KHZ, 3 IS 8 KHZ)
   F)CHAR  8   : CHANNEL TYPE (0 IS NO FILTER, 1 IS FILTER)
   G)CHAR  9   : NOISE (0 IS NO NOISE, P IS +DB NOISE, N IS - DB NOISE)
   H)CHAR 10   : ALWAYS A "."
   I)CHARS11-13: FILE TYPE:
                  1)DS1 IS DIGITIZED SPEECH
                  2)CSY IS COMPRESSED SPEECH (Y IS FRAME PERIOD)
                  3)XYY IS PITCH TRACK(YY IS FRAME PERIOD AND X
                           IS PITCH ALGORITHM:
                           1)CORRELATION           4)HARMONIC
                           2)4-BIT CORRELATION     5)GOLD-RABINER
                           3)2-BIT CORRELATION     6)CEPSTRAL)
                  4)WDB IS NOISE DESCRIPTION (W IS TYPE & DB IS S/N)

PART A:
-------
SPKRS: 1-29, 42-52
DATE: LATE JULY THRU LATE AUGUST, 1978
LOCATION: SC BLDG SPEECH LAB, TEXAS INSTRUMENTS, DALLAS, TEXAS
RECORDER TYPE: TEAC A-4010 (AU OR SL?); 1/4 TRACK, 7 1/2 IPS
DIGITIZED DIRECTLY USING: 980 AIDS: NO    ; VAX AIDS: NO

PART B:
-------

SPKRS: 38
DATE: 1/19/79
LOCATION: SC BLDG SPEECH LAB, TEXAS INSTRUMENTS, DALLAS, TEXAS
RECORDER TYPE: TEAC A-4010 (AU OR SL?); 1/4 TRACK, 7 1/2 IPS
DIGITIZED DIRECTLY USING: 980 AIDS: NO    ; VAX AIDS: NO

PART C:
-------

SPKRS: 30-37, 39-41, 53-58
DATE: OCTOBER-DECEMBER, 1979
LOCATION: HILLCREST SPEECH LAB, TEXAS INSTRUMENTS, DALLAS,TX.
RECORDER TYPE: NO ANALOG TAPE RECORDINGS MADE.
DIGITIZED DIRECTLY USING: 980 AIDS: NO    ; VAX AIDS: YES


TEXT FOR DATA BASE     PITCH
```

MARY HAD A LITTLE LAMB WHOSE FLFECE WAS WHITE AS SNOW. (WD1)
VERY FEW ANGELS ARE ALWAYS WISE AND PURE. (WD2)
THE TROUBLE WITH SWIMMING IS THAT YOU CAN DROWN. (WD3)
NIXON WAS TAKEN TO MOSCOW BY KISSINGER'S AIDE. (WD4)
WHICH TEA PARTY DID BAKER GO TO? (WD5)
AN EXAMPLE OF ONE OF THE BOY'S PERSONAL POINTS IS THE THINNESS OF HIS HANDS. (WD6
A GREAT FUTURE IS ALWAYS PROVIDED THE STUDENT OF MUSIC. (WD7)
IMPORTANT QUESTIONS WERE DRAGGED FROM THE SUBJECT throughout THRUOUT THE MONTHS OF THE TRIAL. (WD8)
ALMOST EVERYTHING INVOLVED MAKING THE CHILD MIND. (WD9)
THE VIEW OF THE PRESENT WILL LARGELY BE REACHED IN THE FOLLOWING CENTURY. (WD10)
THE WIFE'S FIGURE HAD ALREADY ADJUSTED BY ITSELF. (WD11)

SPEAKER DIRECTORY FOR DATA BASE PITCH

	FILE FORMAT	SPEAKERS NAME	SEX M/F	AGE	EDU LEV	PAR BRN USA M	F	TIME IN AREA	VCL TRK PRB	BRTHPLCE
1.	RLDXM31XX.DS1	BOB DAVIS	M	35	19	Y	Y	2	N	ILLINOIS
2.	CJCXM31XX.DS1	CRAIG CATO	M	24	16	Y	Y	2	N	MINNESOTA
3.	BMHXF31XX.DS1	BARBARA HYDRICK	F	33	16	Y	Y	7	N	ALABAMA
4.	CWCXM31XX.DS1	C.W. CLARK	M	30	15	Y	Y	7	N	MISSOURI
5.	JLSXM41XX.DS1	JIM STANFORD	M	47	16	Y	Y	47	N	TEXAS
6.	TJKXM31XX.DS1	TOM KECK	M	28	16	Y	Y	6	N	OKLAHOMA
7.	RGLXM31XX.DS1	GARY LEONARD	M	34	20	Y	Y	7	N	OKLAHOMA
8.	EFGXF31XX.DS1	FREDE GEDDE	F	39	12	N	N	21	N	AUSTRIA
9.	GRDXM31XX.DS1	GEORGE DODDINGTON	M	36	20	N	Y	8	N	FLORIDA
10.	JCLXM31XX.DS1	JOHN LINN	M	32	20	Y	Y	5	N	WASHINGTO
11.	RHWXM31XX.DS1	RICHARD WIGGINS	M	36	20	Y	Y	2	N	LOUISIANA
12.	JLHXM41XX.DS1	JIM HOLMAN	M	50	12	Y	Y	8	N	TEXAS
13.	RNSXM31XX.DS1	BOB SHURTLEFF	M	34	20	Y	Y	7	N	TEXAS
14.	KABXM31XX.DS1	KEITH BLANTON	M	22	17	Y	Y	0	N	TEXAS
15.	REHXM31XX.DS1	GENE HELMS	M	27	17	Y	Y	5	N	ILLINOIS
16.	LFCXF41XX.DS1	LILLIAN CODY	F ~~M~~	59	12	Y	Y	40	N	TEXAS
17.	ALKXF41XX.DS1	LOUISE KLAVITER	F	40	12	Y	Y	10	N	ILLINOIS
18.	DKDXF41XX.DS1	DONNA DUNAWAY	F	43	20	Y	Y		N	TEXAS
19.	ADCXF11XX.DS1	ALISHA CLARK	F	8	2	Y	Y	7	N	TEXAS
20.	SRDXF11XX.DS1	SUZANNE DAVIS	F	11	6	Y	Y	2	N	GEORGIA
21.	MJDXF11XX.DS1	MICHELLE DAVIS	F	8	2	Y	Y	2	N	PENNA.
22.	DRDXM11XX.DS1	DAVID DAVIS	M	6	0	Y	Y	2	Y	PENNA.
23.	SALXM11XX.DS1	SCOTT LEONARD	M	8	2	Y	Y	7	N	MISSOURI
24.	RHWXM11XX.DS1	RICKY WIGGINS	M	11	5	Y	Y	2	N	MASSACHUS
25.	CHWXM11XX.DS1	CHRISTOPHER WIGGINS	M	7	1	Y	Y	2	N	MASSACHUS
26.	ALWXF21XX.DS1	ALLISON WIGGINS	F	13	7	Y	Y	2	N	MARYLAND
27.	LASXF11XX.DS1	LAURA SECREST	F	8	2	Y	Y	8	N	TEXAS
28.	CAWXF31XX.DS1	CAROLYN WIGGINS	F	36	12	Y	Y	2	N	LOUISIANA
29.	PGSXF41XX.DS1	PAT SECREST	F	41	15	Y	Y	12	N	IOWA
30.	OEAXM41XX.DS1	GENE ADAMS	M	48	12	Y	Y	23	N	MISSOURI
31.	ABDXM51XX.DS1	ART DODDINGTON	M	73	8	Y	Y	0	N	MICHIGAN
32.	BJWXM41XX.DS1	BILLY WINN	M	42	12	Y	Y	10	N	OKLAHOMA
33.	CTGXM51XX.DS1	CAPTAIN GILLIAM	M	73					N	TEXAS
34.	EHDXF51XX.DS1	ELNA DODDINGTON	F	80	8	N	N	0	N	ENGLAND
35.	HEMXM41XX.DS1	HUGH METZLER	M	58	14	Y	Y	10	N	KANSAS
36.	PJMXF41XX.DS1	PAT MOORE	F	49	12	Y	Y	40	N	TEXAS
37.	EENXM51XX.DS1	DEAC NYSTROM	M	70	13	N	N	29	N	MINNESOTA
38.	LESXF51XX.DS1	LOIS SECREST	F	75	14	Y	Y	0	Y	IOWA
39.	HLHXF11XX.DS1	HEATHER HURT	F	9	4	Y	Y	9	N	TEXAS

40.	MNNXF51XX.DS1	MYRT NYSTROM	F	70	12	Y	N	29	N	CANADA
41.	RASXM11XX.DS1	ROBERT SHAW	M	10	5	Y	Y	10	N	TEXAS
42.	CJLXF31XX.DS1	JANICE LEONARD	F	35	17	Y	Y	7	N	OKLAHOMA
43.	NLDXF21XX.DS1	NICOLLE DAVIS	F	14	9	Y	Y	2	N	ILLINOIS
44.	LJDXF31XX.DS1	LAURA DAVIS	F	32	13	Y	Y	2	N	OHIO
45.	SXSXF21XX.DS1	SUSAN SCHAFER	F	13	7				N	
46.	HXWXM21XX.DS1	HARRY WINTERS	M	14	8				N	
47.	JXHXM21XX.DS1	JEFF HOPPER	M	14	8				N	
48.	SXCXM21XX.DS1	SCOTT CHEAIRS	M	14	8				N	
49.	SGGXF21XX.DS1	SONYA GOSSETT	F	13	7	Y	Y		N	
50.	SXPXM21XX.DS1	SHANNON PENCE	M	14	8	Y	Y		N	TEXAS
51.	JXMXM21XX.DS1	JIM MOORE	M	15	9	Y	Y		N	CALIF
52.	KXMXF21XX.DS1	KATHY MOORE	F	18	12	Y	Y		N	CALIF
53.	MACXF51XX.DS1	MARTINA CUNNINGHAM	F	70	16				N	WISCONSIN
54.	HDSXM51XX.DS1	H.D. SCOTT	M	84	8	N	N	8	N	ILLINOIS
55.	OEHXM51XX.DS1	ORE HUTSON	M	78	12	Y	Y	3	N	LOUISIANA
56.	NSSXF51XX.DS1	NAOMI SLOAT	F	74	12	Y	Y	2	N	FLORIDA
57.	HOPXM51XX.DS1	HUGH PENNEL	M	87		Y	Y		N	MISSOURI
58.	MBPXF51XX.DS1	MILDRED PENNEL	F	82		Y	Y		N	MISSOURI

APPENDIX B

OPTIMIZED CORRELATION PITCH TRACKER FOR SPEECH SYSTEMS APPLICATIONS

Appendix B

OPTIMIZED CORRELATION PITCH TRACKER FOR SPEECH SYSTEMS APPLICATIONS

George Doddington and Bruce Secrest

Introduction

A pitch extraction algorithm is described which utilizes a segment of speech. containing several pitch frames. The decision as to the pitch period and voicing for a given frame within the segment is deferred to the end of the segment. This helps overcome anomalies in the vocal cord vibrations within the segment and also makes the method robust for speech imbedded in moderate levels of noise.

The algorithm consists of three basis parts. First, a correlation technique is used to obtain the periodicities in the speech to be used as candidate pitch periods. Next, a dynamic programming algorithm using these candidate pitch periods is used to preserve the continuity of the pitch track. Then, as a last step, pattern matching with the correlation values from the optimal track is used to obtain the voiced/unvoiced decision. The three basic steps of the pitch extraction algorithm will be discussed in the next three sections.

Candidate Pitch Periods

The candidate pitch periods for each pitch frame are obtained by using a normalized correlation technique. Since the frame of data to be analyzed is imbedded in a segment of speech, a forward correlation into new speech is accomplished as well as a reverse correlation into old speech. This allows better candidate pitch periods in regions of transition such as from nasals to vowels.

Given a frame of speech data consisting of N samples (typically 10 milliseconds), the first M samples of this frame (typically 10 milliseconds), W_f, are used as a sliding window for the forward correlation and the last M samples of the frame, W_r, are used in their reverse correlation. The normalized cross-correlation, $R_f(K)$, between the sliding window, W_f, and M speech samples beginning at the K^{th} sample is used for the forward correlation and is defined as

$$R_f(K) = \left[\sum_{m=1}^{M} X(m)X(K+m-1)\right] \Bigg/ \left[\left(\sum_{m=1}^{M} X^2(m)\right)\left(\sum_{m=1}^{M} X^2(K+m-1)\right)\right]^{\frac{1}{2}} \qquad (1)$$

$$(K_{min} \le K \le K_{max})$$

where x(i) is the value of the i^{th} speech sample and K_{min} (= 2 milliseconds) and K_{max} (= 25 milliseconds) correspond to the minimum and maximum pitch periods to be considered, respectively. Similarly, the normalized cross-correlation, $R_r(K)$, between the sliding window, W_r, and M speech samples earlier in time starting at the $(M - K_{min})$ samples is used for the reverse correlation, i.e.:

$$R_r(K) = \left[\sum_{m=1}^{M} X(N-m+1)X(N-m-K+1)\right] \Bigg/ \left[\left(\sum_{m=}^{M} X^2(N-m+1)\right)\left(\sum X^2(N-m-K+1)\right)\right]^{\frac{1}{2}} \qquad (2)$$

$$(K_{min} \le K \le K_{max})$$

Note that $| R_f(K) | \le 1$ and $| R_r(K) | \le 1$ because of the normalization.

Once the $R_f(K)$ and $R_r(K)$ have been computed from equations (1) and (2), a set of candidate pitch periods, S, is obtained by picking those values of K for which $R_f(K)$ and $R_r(K)$ attain a maximum or peak ($\tilde{R}_f(K)$ and $\tilde{R}_r(K)$. These peaks must be such that $\tilde{R}_f(K) \geq .5$ and also $\tilde{R}_f(K)$ must be 1.3 times the previous minimum or valley in the function, with similar constraints on $\tilde{R}_r(K)$.

The set of candidate pitch periods, S, is enlarged by adding the half pitch period, K/2, for all K S such that K K_D; where K_D is a fixed value (= 8 milliseconds). Also added to the set S is the unvoiced candidate or no pitch period. Thus, if no maximum or peak of either $R_f(K)$ and $R_r(K)$ satisfy the above constraints, the set S contains only the unvoiced candidate.

Optimal Pitch Track

Given the set S of candidate pitch periods for each pitch frame in the segment of speech, it is desired to extract a pitch period for each frame such that the pitch track is continuous across the entire segment, contains the pitch periods with the higher cross-correlation values, and minimizes pitch period doubling. A dynamic programming algorithm is used to achieve these goals.

The dynamic programming algorithm consists of T trajectories (T = 4) or tracks through each pitch frame in the segment of speech. At each pitch frame, i, the j^{th} trajectory consists of a pitch period, K_i^j, the value of a cumulative penalty, P_i^j, and a back pointer, B_i^j, to that trajectory in the previous frame resulting in the minimum cumulative penalty.

To extend the trajectories to the current $(i + 1)^{st}$ frame, each element, $K_{i + 1}$, of the set S of candidate pitch periods for the current frame is compared with all T trajectories of the previous frame. This comparison consists of assessing a penalty in going from the i^{th} frame to the $(i + 1)^{st}$ frame. The cumulative penalty at the $(i + 1)^{st}$ frame using the j^{th} trajectory of the i^{th} frame is given as:

$$P^j_{i+1}(K_{i+1}) = P^j_i + E_{i+1} \left[1-R(K_{i+1}) + \beta K_{i+1} + \alpha \,|K_{i+1}-K^j_i|\right] \quad ; \qquad (3)$$

where P^j_{i+1} (K_{i+1}) is the cumulative penalty for the j^{th} trajectory at the $(i + 1)^{st}$ frame using the candidate pitch period K_{i+1} from set S at frame $(i + 1)$; P^j_i is the cumulative penalty for the j^{th} trajectory at the i^{th} frame; E_{i+1} is the RMS energy in the sliding window, W_f, at the i^{th} frame, K^j_i, are also extended into the $(i + 1)^{st}$ frame with a constant penalty being added. $(P^j_{i+1}(K_i) = P^j_i\ (K_i) + 1 - .5 + (.003)(40))$.

At any frame, the set of cumulative penalties obtained by the method described above is search to find the T minimum cumulative penalties. These T trajectories are then saved for that frame to be used in extending to the next frame.

Another way to look at the dynamic programming approach used in the algorithm is to say that in order to maintain pitch track continuity across the speech segment, several frames (at least four) are analyzed before deciding upon the first frame. At each frame, every pitch candidate is compared to the retained pitch candidates of the previous frame (only four pitch candidates are retained for each frame). Each comparison results in a cumulative penalty and there will be a smallest penalty for each of the candidates in the new frame corresponding to a comparison to one of the retained pitch candidates of the previous frame. In addition, each pitch candidate of the previous frame is also a candidate in the new frame with a fixed increasein cumulative penalty. When the lowest cumulative penalty has been calculated for all new candidates, the four with the lowest cumulative penalties are retained, along with their cumulative penalties, correlation peak values and back pointers. The back pointer of a pitch candidate indicates which candidate of the previous frame corresponds to its cumulative penalty. Likewise that candidate in the previous frame has a back pointer identifying another candidate in the frame before it, etc. Thus the back pointers define a trajectory which has the associated cumulative penalty of the last analyzed frame. The cumulative penalty at the $(i + 1)^{st}$ frame of the j^{th} trajectory is given by equation (3). After the four candidates and associated parameters of a pitch frame have been obtained, that trajectory with the lowest cumulative penalty, P_i^n is selected as correct. It is traced backward m frames (at least four) to find the pitch value, $K_{i - m}^n$, identified as the pitch during the $(i - m)^{th}$ frame.

At the end of the segment of speech, the T trajectories in the last frame are searched for the minimum cumulative penalty. The trajectory describe̵d by the backpointers is called the optimal pitch track for that segment of speech.

Voiced/Unvoiced Decision

Given the optimal pitch track from the dynamic programming algorithm, the correlation values of the pitch periods of this optimal path are scanned to make a voiced/unvoiced decision at each frame.

The scanning patterns are meant to span L (= 4) frames of the segment of speech, which corresponds to L time periods. The motivation for the scanning patterns is that when the speech is unvoiced, the correlation values should be high. Note that the correlation values for the optimal pitch track will vary from .5 to 1.0. In determining changes from voiced to unvoiced speech, the correlation values would expect to decrease from a high value to a low value in a few time frames, and vice versa for the unvoiced to voiced transition. With this in mind, four-point scanning pattern vectors might look like (L = 4):

$$\begin{array}{lll}
\underline{P}_{UV} = \{.5, .5, .5, .5\} & \text{(Unvoiced)} & (4) \\
\underline{P}_{V} = \{.9, .9, .9, .9\} & \text{(Voiced)} & \\
\underline{P}_{VUV} = \{.8, .8, .8, .8\} & \text{(Voiced to unvoiced transition)} & \\
\underline{P}_{UVV} = \{.5, .5, .5, .5\} & \text{(Unvoiced to voiced transition)} &
\end{array}$$

Four errors are determined for each frame of speech by centering the scanning patterns on the second element of the vector and computing a squared error between the scanning pattern and the correlation values of the optimal pitch track, i.e.:

$$E_I^i = \sum_{j=1}^{4} \left[R(K^o_{i-2+j}) - P_I^{\,j}\right]^2 \qquad I = UV,\ V,\ VUV,\ UVV \tag{5}$$

where E_I^i is the scanning error for the i^{th} frame for one of the four scanning patterns: $\underline{P}_{uv}$, $\underline{P}_v$, $\underline{P}_{vuv}$, and $\underline{P}_{uvv}$; $R\ (K_i^o)$ is the correlation value for the pitch period at the i^{th} frame contained in the optimal pitch track; and P_I^i is the i^{th} element of the I^{th} scanning pattern.

The voiced/unvoiced decision is made by comparing the scanning errors against fixed thresholds. If the sequence is $\left\{\begin{matrix}\text{voiced}\\ \text{unvoiced}\end{matrix}\right\}$ at the $(i - 1)^{st}$ frame, then the $\left\{\begin{matrix}E_v^i\\ E_{uv}^i\end{matrix}\right\}$ scanning error is compared with a fixed threshold (= .4). If this error is less than the threshold, the i^{th} frame is changed to $\left\{\begin{matrix}\text{unvoiced}\\ \text{voiced}\end{matrix}\right\}$. If this error is also larger than the threshold, the decision is deferred. The above strategy is continued until a frame either is confirmed as being the same, i.e. $\left\{\begin{matrix}\text{voiced}\\ \text{unvoiced}\end{matrix}\right\}$ then any intermediate frames which were deferred are made $\left\{\begin{matrix}\text{voiced}\\ \text{unvoiced}\end{matrix}\right\}$. However, if the voicing decision has changed, i.e. $\left\{\begin{matrix}\text{unvoiced}\\ \text{voiced}\end{matrix}\right\}$ and there are intermediate frames which are unresolved, then the scanning pattern errors $\left\{\begin{matrix}E_{vuv}^i\\ E_{uvv}^i\end{matrix}\right\}$ are searched for their minimum value at these intermediate frames and the transition point is set at this minimum point.